GRANT
AF EOAR 64-54

ILRS - T1i, 650630
June, 1965

FINAL SCIENTIFIC REPORT

AUTOMATIC ENGLISH SENTENCE ANALYSIS

IDAMI LANGUAGE RESEARCH SECTION

Istituto di Documentazion[illegible]
Piazza S. Ale[illegible]

[illegible]'Associazione Meccanica Italiana
M I L A N, Italy

GRANT AF EOAR 64-54 30 June 1965

FINAL SCIENTIFIC REPORT

AUTOMATIC ENGLISH SENTENCE ANALYSIS

IDAMI Language Research Section

Istituto di Documentazione dell'Associazione Meccanica Italiana

Milan, Italy

The research reported in this document has been sponsored by

THE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

under Grant AF EOAR 64-54 through the European Office of Aerospace Research (OAR), United States Air Force.

A B S T R A C T

The research summarised in this report was carried out by two groups working collaterally under GRANT AF EOAR 64-54; the report, therefore, consists of two parts.

Part I: The LINGUISTICS Group describes work on the Multistore Procedure for analysis of English sentences. Previous reports are given as abstracts and subsequent developments are described.

Machine-economies have been introduced in various aspects of the procedure; the general table of correlators has been refined; a corpus of texts has been key-punched and processed to provide data for analysis of explicit correlators; this analysis is partly completed. The procedure for reclassification of intermediate products is described. Five appendices illustrate the present state of the work.

Part II: The MATHEMATICS Group reports work on aspects of the structure of language. Considering the sentence as codification of thought, and thought as a complex of mental items between which certain relations hold, the report considers the source of these mental items, under three headings: 1) sense-perceptions, 2) reflection, 3) inter-personal communication. The structure of language is found to be characterised by the way in which relations between mental items are codified.

PREFATORY REMARKS

The scope of Grant AF EOAR 64-54, under which work on the project started in March 1964, was extended by an Amendment in December 1964 to include specific mathematical research. As a result of this the team was enlarged and divided into two separate groups, one concerned mainly with the linguistic aspects of Automatic English Sentence Analysis, the other with the study of mathematical applications to this field.

The amendment also changed the original terminal date of the Grant from March 1966 to 30 June 1965; the working period for the linguistics group was thus shortened by nine months, while the newly instituted mathematics group was allotted a working period of seven months.

In April 1965 a new Grant was stipulated (AF EOAR 65-76) to provide support for the continuation of the original linguistics research from July 1965 to October 1966, and for further mathematical research from July 1965 to March 1966.

The two groups, although working collaterally, are concerned with different problems and different methods and this report, therefore, is divided into two parts:

Part I - summing up the linguistics research during the period 1 March 1964 to 30 June 1965,

Part II - summing up the mathematics research during the period 1 December 1964 to 30 June 1965.

TABLE OF CONTENTS

Title of Report

Abstract

Prefatory Remarks

Table of Contents

page

PART I

Research of the Linguistics Group

Acknowledgments 1

Progress of Work 3

RESEARCH REPORT

Background 5

Summary of Report T3 5

Division of Tasks 6

1) Limited Vocabulary 9

2) General Table of Correlators 11

3) Corpus and Concordance Program 13

4) Analysis of Explicit Correlators 15

Summary of Report T9 15

Levels of the Analysis 15

Difficulties of the Analysis 16

5) Assignation of Correlation Indices 19

6) The Multistore Procedure 21

Summary of Report T10 23

Recent Advances 23

7) Reclassification Procedure 27

Program of Work (next Grant-period) 37

Reports and Publications 39

APPENDICES

No. I: Vocabulary of the Project I/1 - I/5

No. II: Table of Correlators II/1 - II/11

No. III: Break-up of the Correlator "BY" (37 analysis sheets)

No. IV: Sample of Index Assignation IV/1 - IV/4

No. V: Sample of Reclassification V/1 - V/3

P A R T II

page

Research of the Mathematics Group

Introduction 1

Fundamental Research already carried out 1

Further Projects 9

Introduction to the Study of Language 10

The Hypothesis of the Mind as a Factory for the Production of Ideas 11

Mental Operations 12

P A R T I : Research of the LINGUISTICS GROUP

Principal Investigator: Ernst v. Glasersfeld

Research Associate: Jehane Burns

Programming Expert: Pier Paolo Pisani

Research Assistant: Brunella Notarmarco

A C K N O W L E D G M E N T S

We thank the Computation Center of the Electronics Division of OLIVETTI, Milan, which put at our disposal their ELEA 9003 for part of our experiments.

We thank also the UNIVAC Division of Remington Rand, Italy, who have kindly offered to allow us the use of a computer for experiments in the coming months.

Some of the publications received during the Grant period, which we have found of special relevance or interest, are listed in the General Bibliography at the end of Part I of this Report. We wish to thank all the research groups who have sent us publications.

We are most grateful also to our Project Monitor, Rowena Swanson, who has sent us material of great interest which would not otherwise have come to our attention.

PROGRESS OF WORK

Progress in the Linguistics Sector was satisfactory with regard to conceptual work, and the testing of the main parts of the sentence analysis procedure by manual implementation on a specially designed display panel proved very successful. Valuable insights were gained in the various problem areas (see sections 2, 4, 5, below). Development of the Multistore Procedure (see section 6, below) has gone a good deal further than expected and the system has reached a higher degree of coherence and sophistication than we had hoped for at the outset. The problem of "Reclassification" (see section 7, below), the technical difficulties of which were impossible to assess before the system had been worked out in detail, has been overcome in principle; although errors in individual index assignation are expected to crop up for quite some time to come, the procedural aspect of the problem was solved without much difficulty.

A major setback in the progress of work was caused by the breakdown of arrangements we had made with private firms for the use of their computers. In one case this was particularly unfortunate, because it involved the loss of three months of programming work, which had to be scrapped as the computer for which the programs were being written was suddenly dismantled and no other machine of the same type could be found. - We had a similar disappointment with another firm, but this time, fortunately, it involved much less loss of work and time.

These setbacks, however, in conjunction with the loss of time caused by the search for other machines and by the inevitable reorganisation and rewriting of programs, have brought about a serious delay in the machine testing

of the system; as a result of this we are unable to include in this report the fully operational machine program of the Multistore Procedure, which we announced in our report of January 1965. The only redeeming feature of this delay is the fact that the procedure which will be programmed during the next research period constitutes an advance on the procedure we expected to have programmed by now.

R E S E A R C H R E P O R T

Background

The theoretical background of the project "Automatic English Sentence Analysis" and the course the research work was to take, were outlined in our Informal Report of May 1964, which was subsequently published in English by the Munich journal Beiträge zur Sprachkunde und Informationsverarbeitung (see "Reports and Publications", below); since neither the guiding principles nor the ultimate aims of the project have undergone noteworthy changes, we here merely include a summary.

Summary of Descriptive Report T3

The purpose of this project is twofold:

a) to elaborate and improve a correlation procedure (cf. Ernst v. Glasersfeld and Jehane Barton Burns "First Draft of an English Input Procedure for Mechanical Translation", Technical Status Report No. 8, Contract AF61(052)-362, Centro di Cibernetica e di Attività Linguistiche, Milan University, February 1962) and to implement it in a computer program by means of a "Multistore" system;

b) to use this essentially syntactic correlation procedure to isolate and determine, within the limits of a restricted vocabulary, "semantic factors" indispensable for the correct analysis (i.e. interpretation of meaning) of English sentences and, subsequently, to employ these semantic factors in the correlation procedure in order to reduce the output of syntactically, but not semantically, acceptable analyses.

The project is based on the theoretical work of Silvio Ceccato (cf. General Bibliography, Ceccato, 1960) which

it intends to extend and adapt through the empirical application to present-day English. A considerable part of the research will be directed towards analysis of "correlators", i.e. on the one hand, the relations the human mind sets up between the individual items in a train of thought, on the other, the linguistic expression of these relations, and their interdependence. A comparative study of correlator expressions is planned for English, Italian, French, German, and, if possible, Russian.

Division of Tasks

The first five months of the Grant period (March-July 1964) were taken up with the systematic organisation of the material that was to serve as basis for the planned research. Work was divided into a number of more or less separate tasks:

1 - Selection of about 100 English "head-words" which, together with their grammatical derivatives, were to constitute the limited vocabulary of the project.

2 - Revision, development, and adaptation of the General Table of Correlators for application to the analysis of English sentences.

3 - Preparation of a corpus of modern English texts and of a concordance program as research aids.

4 - Analysis of the explicit correlators contained in the limited vocabulary.

5 - Assignation of correlation indices to the words of the limited vocabulary.

6 - Development and programming of the "Multistore" Procedure for syntactic analysis.

7 - Elaboration of rules for the assignation of correlation indices (reclassification) to individual

correlations, and implementation of these rules in a reclassification procedure.

In reporting the research results obtained up to the present date, we shall maintain this division and treat the seven tasks separately.

1 - Limited Vocabulary

The main objective in the compilation of the vocabulary was that the words, although few in number, should allow the formulation of sentences showing the greatest possible variety of syntactic construction. The criterion for including a word, therefore, was not its frequency rating in a statistical examination of modern English usage, but its syntactic versatility.

Since the project, at this first stage, is not concerned with the analysis of sentences containing punctuation marks, such words or word functions as can occur only in constructions with obligatory commas, colons, etc., were deliberately excluded.

The vocabulary, as it stands at present, forms Appendix No. I of this report. For the sake of convenience it has been arranged in the more or less traditional grammatical classes.

2 - General Table of Correlators

An exposition of the Correlation Theory of the Italian Operational School, on which our project is based, can be found in our Report T10, pp. 5-16;(see the section "Publications"). We began by revising and adapting specifically for English the correlation table we had previously compiled at the Centro di Cibernetica (see Bibliography, Ceccato, 1963).

The correlation table we are presenting in this report (Appendix No. II) still falls short of the "ideal" we are aiming at. In more than one place it reflects compromises which were made deliberately (e.g. correlations 381-385) in order to be able to use the table in the preliminary experimental version of the Multistore procedure, which does not, as yet, allow for correlational discontinuities (i.e. for the correlation of words, or word combinations, that are not contiguous in the sentence to be analysed).

The present table still contains groups of correlations which have not yet reached the degree of differentiation that will, ultimately, be required (e.g. correlations 401, 411, and 441-487).

Finally, as well as the shortcomings we know of, there probably are others of which we have not yet become aware. Nevertheless the table, as it now stands, has been used in a certain number of manual simulations of sentence analysis and has yielded very satisfactory results (see the example of procedure in Report T10, pp. 65-77).

3 - Corpus of Texts and Concordance Program

As an aid in compiling the table of implicit correlators and in the analysis of explicit correlators, 24,000 words of English texts were key punched and processed with a concordance program. The texts were chosen from works by William James, Bertrand Russel, Aldous Huxley, and Paul de Kruif (for the variety of their prose and of the generally scientific subjects treated) and from D.H.Lawrence's "Twilight in Italy" and James Joyce's "Dubliners" (because literary prose provides a more varied and complete survey of syntactic constructions than does scientific writing); to facilitate parallel research in Italian, IDAMI provided for the key punching and processing of the corresponding Italian translations and it was agreed to add to the English corpus a specimen of Colquhoun's translation of Alessandro Manzoni's classic "I Promessi Sposi".

The print-out of concordances, apart from providing useful data about word-frequencies, provided a valuable stock of examples of correlator occurrences, which, together with the examples found in Henry C. Wyld's "Universal English Dictionary", Webster's "New International Dictionary", and the "Concise Oxford English Dictionary" formed the basis of our first-level correlator analyses. When the break-up of prepositions and conjunctions (cf. section 4 of this report) is finished, we shall test the results against this corpus, which, although too limited to prove the general validity of our analyses, will at least help us to eliminate the grosser kinds of error.

4 - Analysis of Explicit Correlators

Our approach to the problem of explicit correlators - or, in general, the problem of isolating and identifying the different relations expressed by prepositions and conjunctions - was outlined in our report of May 1964 (T3, pp. 3-6) and more fully treated in the subsequent report T9 of January 1965 (see the section "Publications", below) of which the following is a summary.

Summary of Report T9

The paper gives a brief account of sentence analysis in terms of a finite set of correlators, and examines the analogies and differences between automatic sentence-analysis and some of the activity of the human mind in understanding discourse. The importance, in sentence analysis and in translation, of the interpretation of words such as prepositions is pointed out, and some of the problems involved in designing an automatic interpretation procedure for prepositions are discussed.

Levels of the Analysis

A first level of analysis was reached before the end of 1964. The 27 explicit correlators contained in the project's vocabulary (cf. Appendix No. I) had been broken up into subsets of relations according to Italian and German correspondences. The average number of relations isolated for each correlator was 15; their distribution, as could be expected, was extremely uneven, ranging from one relation in the case of the correlator "though" to a maximum of nearly fifty relations in the case of the correlator "of". As a sample of this first stage of analysis the break-up of the correlator "about" was attached as Appendix No.3 to Administrative Report No. 1 (August 1964).

This part of the research has been continued. At present the individual relations are being tested for French correspondences. As a sample of this second level of analysis we are attaching the break-up of the correlator "by" as Appendix No. III to this report.

Difficulties of the Analysis

One of the difficulties in the analysis of explicit correlators is that of establishing what, precisely, is the distinguishing feature of a set of English correlations that assures homogeneous output in another language. A good example of this difficulty is the use of the correlator "by" in expressions such as "we'll get the house ready by Friday" (cf. Appendix No. III, I_r's 051 and 052). To an English speaker the relation expressed by "by" in this sentence seems identical with the one the "by" expresses in "we"ll get the house ready by the summer". Translation of the two sentences into Italian confirms the impression of identity, since in both cases the "by" can be rendered by the Italian correlator "entro". Translating the two sentences into German, however, we find that "by Friday" becomes "bis Freitag", while "by the summer" becomes "bis zum Sommer". - If the German informant is aware of one's preoccupation with homogenous output, he will probably try to skirt round the difficulty by translating "vor" in both cases; but this would be correct only if the English sentences were made with "before", which sets up a somewhat different relation. - So one presses on to establish why the German correlator "bis", in this context, goes well with the days of the week and the months of the year, but has to be supplemented with the correlator "zu" when the point of reference happens to be a season; and eventually one comes to realise that the operative difference is something like this: in the German system of relating words to experience

(the "semantic universe" of German, if one wants to sound philosophical) the names of days and of months are, as such, regarded as purely temporal markers, while the names of seasons tend to refer also to the characteristics habitually associated with those temporal indications. In other words, there is something in the designata of "Winter", "Sommer", "Herbst", etc. which justifies, or even enforces, their being treated like <u>events</u> such as are designated by "thunderstorm", "high tide", "full moon", and the like. An English speaker who wanted to express that his friend Mary will have returned to him when the full moon next rises, would scarcely say "by the full moon" (*); a German, however, can say "bis zum Vollmond ist Marie zurück", just as he can say "bis zum Frühling ist Marie zurück". It is on the strength of this difference that we assign to "by", whenever its second correlatum is the name of a season, a relation index I_r that is different from the one assigned to it when its second correlatum is another kind of temporal specification.

We have no doubt that for all, or at least for most, of these discrepancies in the use of prepositions in different natural languages one can discover theoretical reasons of this sort; but, although we are rather interested in this particular aspect, it is not our first concern. Our immediate objective in this kind of correlator analysis is to arrive at a system of I_r indices capable of reducing ambiguity and, consequently, multiple output.

This purpose entails the formulation of working criteria which accurately pin down the operative features of each relation which needs to be distinguished, so that we can use them to determine classes of correlata <u>accept-</u>

(*) He might say "by the time the full moon...", thus specifically supplying the temporal aspect required by the relation he wants to set up.

able as first and as second correlatum of each I_r. In this way, any correlation accepted under the heading of a particular I_c will give riss not to the hole gamut of I_r's of that I_c but only to a restricted set - ideally, only to one I_r.

The samples of I_r-definitions given in Appendix No. III are still far from satisfactory. As yet, no attempt has been made to organise and homogenise the working-criteria of the relations and the descriptions of the correlata they involve. We firmly believe that it would be detrimental to the ultimate results in this sector of research, if the attempt to bring the diverse definitions into a coherent system were made before these definitions have been tested empirically on a larger scale. Past experience - and not only ours - has shown how powerful a temptation it is to adjust conclusions drawn from usage to the preconceived requirements of theoretical assumptions.

When all the correlators we are concerned with are brought to the second level of analysis, we shall index the words of our vocabulary according to the provisional I_r-code (*). This will be a first test for the validity of the break-up. When this test - and, if possible, one with a much richer vocabulary - has been completed, and when its results are fed back into the break-up, we shall have the necessary basis to set up a "General Table of Relations", i.e. an ordered list of the relations that can be expressed by means of the explicit correlators contained in the vocabulary.

(*) It should be noted that the I_r-numbers we have so far assigned to individual relations (cf. Appendix No.III) are provisional and do not reflect any logical order; they were assigned in the order in which the relations cropped up during analysis.

5 - Assignation of Correlation Indices to Words

This task has been completed: all the words of our vocabulary have been assigned their individual correlation indices and the strings of indices referring to implicit correlators have been corrected and are at present in accordance with the actual version of the general table of these correlators; the strings of indices referring to explicit correlators do not yet reflect the analyses carried out on these correlators, because they will be finished only towards the end of 1965 (cf. section No. 4, "Explicit Correlator Analysis").

Samples of word-cards with their strings of correlation indices (I_c) are given in Appendix No. IV to this report.

6 - The Multistore Procedure

The Multistore Procedure is a system of automatic sentence analysis designed for application to sentences that consist of words listed in a pre-established vocabulary.

The procedure is based on the Correlation Theory first formulated by Silvio Ceccato and the Italian Operational School. This theory supplies a substitute for traditional syntax and grammar, insofar as it describes the concatenation of words in any given sentence as the result of correlations rather than of syntactic functions and of the interplay of the grammatical classes of the occurring words (cf. Introduction to Correlation Theory, Report T10).

A finite set of individually characterisable correlators is derived by empirical research on texts (see "General Table of Correlators", section 2, above, and Appendix No.II). The words of the vocabulary are then examined as to their correlational possibilities in the light of the listed correlators. These correlational possibilities (i.e. possibilities of entering into correlation with other words, by means of a correlator) are recorded on the word-cards in the form of correlation indices 'I_c' (see "Assignation of Correlation Indices", section 7, below).

Given this disposition of data, sentence analysis becomes a question of examining which correlations are possible on the strength of the I_c's borne by the words that compose the sentence to be analysed.

While previous implementations of correlational analysis were conceived as a continuous reiteration of index comparisons (in order to ascertain the compatibilities, within the given sentence, of single words with single words, single words with word combinations, and, finally word combinations with word combinations), the Multistore system, owing to its particular internal structure, elim-

inates this cumbersome succession of comparisons and identifies the existing compatibilities in one organic process based on the pigeonholing of the relevant data.

Its central organ is a memory divided into 'columns', each of which is reserved a priori for one of the correlators of the general table. As the I_c's of each word-card (corresponding to a word of the given sentence) are read, for each of them a 'marker' is inserted in the corresponding column of the Multistore; the I_c, that is, <u>is</u> an address to a specific column; the presence of a marker in a Multistore column thus indicates one particular correlational possibility of the word to which the marker belongs; and as soon as a compatible ('complementary') marker finds its way into the same column, the very fact of its insertion shows that the particular correlation represented by that column is actually possible in the given sentence. We say "finds its way into the column", because entry of the complementary marker can be made to depend on a variety of conditions derived from the rules of word-order pertaining to the input language, from specific rules pertaining to the particular correlator, and, finally, from the particular position (e.g. first or last) the word to which the marker belongs happens to have in the sentence at hand.

The ease with which such rules can be incorporated into a coherent system, and the flexibility of application that can be achieved, is the second main asset of the Multistore procedure. - In short, the system, on the one hand, does away with one-to-one index comparisons, and, on the other, reduces the correlational possibilities of a given word (i.e. the possibilities that word may have in <u>any</u> sentence) to those possibilities which the word has in its particular place in the sentence at hand.

A full description of the Multistore system can be found

in our Report T10 of January 1965.

Summary of Informal Report T10, January 1965

After a short introduction to correlation theory, the paper explains and describes a new procedure for correlational analysis for single sentences in English. It shows how grammatical, syntactic, and semantic data can be handled in one coherent program by means of a specially organised memory which we have called the 'Multistore'. The system eliminates the need for comparisons of indices and ascertains the compatibility of elements (in this case, words and phrases) by inserting the indices that characterise each element into pre-established individual 'columns', each of which has its own fixed combination rules based on the rules of word-order in current English. The system yields a complete listing of the operative correlations (roughly corresponding to syntactic connections) and their hierarchical structure in a given sentence. - Summary flow charts and an example of the procedure are added to the text.

Recent Advances

Since the publication of this report, the procedure has been refined in several ways.

a) The system of restriction rules (i.e. rules which restrict the number of correlation indices entering into the procedure to those which are relevant in the given sentence) cuts out all those indices which reflect correlational possibilities assigned to a word considered in itself before its particular position in the sentence has been taken into account but which cease to be realisable given the particular position the word has in the sentence at hand.

b) Special subroutines for the blocking and unblocking of certain Multistore-columns (approximately 30%) have been devised and have resulted in a considerable further economy with regard to the insertion of superfluous 'markers' into the columns of the Multistore.

These subroutines are based on the fact that certain correlations can occur in a given sentence only if certain words occur in the sentence. Correlations 201 N or M, for instance, can become operative only if the word "am" is present in the sentence that is to be analysed. Similar conditions can be formulated for approximately 75 other correlations.

The 'Preliminary Pass' (cf. Report T10, pp.24-27) was therefore expanded to comprise a check for the presence (in the sentence at hand) of a certain number of individually listed words. If one or more of these words are found to be present, the correlation columns in the Multistore corresponding to the particular indices conditioned by the presence of these words, are opened for operations; the columns depending on specific words which are <u>not</u> present in the given sentence, on the other hand, remain blocked, and, for this sentence, the procedure works as though these correlations did not exist.

This further restriction of the general possibilities of a word to possibilities that are actually feasible in the given sentence, has greatly reduced the insertion of 'markers' and thus the amount of data that has to be processed in the Multistore.

c) A number of further restriction rules (again restricting the insertion of 'markers' into the Multi-

store) were devised on the strength of certain rather obvious but useful observations regarding the effect of the presence (in a given sentence) of explicit correlators on the implicit correlational possibilities of the words and word combinations immediately before and immediately after the explicit correlator.

Thus, for instance, nearly all the special rules formulated for the first word of a sentence can be applied to the second word of the sentence if the first word happens to be an explicit correlator. (e.g. If a sentence begins with "because", the second word - even if this be one of its general possibilities - cannot possibly have the function of verb in a non-inverted 'subject-verb' correlation; nor can it be the second correlatum in any other non-inverted correlation that does not concern the particular explicit correlator "because".)

Similar restrictions can be introduced if the last word of the given sentence happens to be an explicit correlator (which, in this position, would not function as correlator, but as correlator adverb); and some restrictions can be applied to the words immediately before and immediately after an explicit correlator wherever it occurs in a given sentence.

d) The reclassification routines (described generically in Report T10, pp. 35-37) have been worked out in detail and incorporated into the Multistore procedure (see section 7, below).

7 - Reclassification

Reclassification is the assignation to complete correlations, produced during the analysis of a sentence, of strings of indices like those which appear on the word-cards. It is these indices which will permit the product to become a correlatum in further correlations, that is to say, they are read and used in just the same way as the I_c's of individual words.

The I_c's of individual words, however, are assigned to each word "by hand" by the linguist. Even though he can draw some of the I_c's from pre-established lists according to formal characteristics of the word, - for example, any present participle must receive certain indices, such as 256 N2 - in many respects he tailors the I_c's to the special possibilities of the word; the different "framing" functions of verbs, for instance (we can say "I want to see you", "I imagine he is here", but not "I imagine to see you", "I want he is here", while the verb "hope" and "expect" have both these two functions) can be implemented once and for all by the assignation of the appropriate and only the appropriate indices to the verbs involved.

These are "rules" which fall within the traditional domain of syntax; but in his assignation of I_c's the linguist can also embody simple semantic data. For instance, he can arrange that the Multistore shall accept the correlation "gloves on" in the sentence "he arrived with gloves on", but not the correlation "reflections on". Thus in a sentence "he began with reflections on ....." the procedure will not wrongly correlate "on" adverbially. Similarly, the linguist can exclude, in a sentence such as "he lamented past opportunities", the analysis which would take "past" as a correlator (as in "I walked past the bank") simply by not giving "opportunities" the I_c of 2nd correlatum

of the correlator "past".

Decisions of this sort, made by the linguist in assigning I_c's to individual words, are ad hoc and do not require the formulation of general rules; the restrictions they impose are, as it were, tacit, and nowhere appear explicitly.

In the assignation of I_c's to products, however, we no longer have to do with individuals, and we can no longer make the assignation direcly "by hand". We have to enable the machine, when it has made a product (i.e. a correlation of two or more words), to assign to it the appropriate I_c's. A product has three variables: its correlator, and its two correlata; and all three influence the choice of I_c's. If we could treat every combination of these three variables as an individual, we would be able to reduce the I_c-strings of the product and exclude many irrelevant possibilities. But this obviously is not practicable even with first-level products. We have to treat products by classes, reducing the number of I_c's as far as possible by means of classes, but taking care to exclude no index which might be relevant to any individual product within a class.

The first variable in classifying products is the correlator. When a product is made and inserted in the product store, it bears on its face, as it were, the number of its correlator; whereas the characteristics of its correlata can only be found by re-examination of their I_c-strings. Further, one of the criteria in the composition of the general list of implicit correlators - in deciding how far they should be subdivided - was that each correlator should be as far as possible homogeneous in the reclassification of its products.(*)

(*) By subdividing correlations further, in terms of classes of their correlata, the process of reclassification of products could of course be further simplified, but at the expense of multiplication of columns and markers in the Multistore. The optimum balance between classification and reclassification is impossible to determine a priori; given the design of the Multistore procedure, however, changes of this kind can be made at any stage.

Reclassification works, then, on the basis of the correlators; that is, each correlator has its individual reclassification procedure, which is called into operation when a product is made by that correlator. This procedure does not consist simply in the assignation of a single I_c-string; the influence of the other two variables of the product, the correlata, can be used to reduce the number of indices to the extent to which this influence can be foreseen by the linguist in terms of <u>classes</u> of first or second correlata; further, it must be possible to make a correlatum's membership or non-membership in one of these classes recognisable to the machine on the basis of the information about the correlatum which is available in the product store.

Instead of a single I_c-string, then, the linguist has to compile for each correlator (*) a list which in most cases will consist of a set of indices which are to be assigned unconditionally (that is to say, which are assigned to <u>every</u> product made with the correlator in question) and of several other sets of indices, each of which is to be assigned to the product in hand only if its correlata fulfill certain conditions. These conditions are of various kinds, but in every case the product can be tested for them by reference to information in the product store.

Apart from the "fixed list" which forms part of the reclassification of nearly every correlation, the conditions for the assignation of an I_c to a product are of the following types:

a) Indices which are assigned if they are present in the I_c-string of the first correlatum. In many implicit correlations the product has obviously, and

---

(*) An N-correlation and its corresponding M-correlation have quite independent reclassifications; the same is true of E and F.

constantly, much more in common with one of its correlata than with the other; the phrase "green string" behaves very similarly to the word "string", quite differently from the word "green"; "to eat omelettes" behaves more like "to eat" than like "omelettes"; "nuts in May" is more like "nuts" than "May".(*) In a case where the product has much in common with its 1st correlatum, there will be an extensive list of reclassification I_c's (the sum of several sets of I_c's appropriate to different classes of 1st correlatum) bearing the condition "if present in I_c-strings of 1st correlatum".

In some cases it is possible to give a general rule "assign all the indices of the 1st correlatum" (e.g. in the reclassification of correlation 462 N); but in most cases not all the I_c's are applicable; for example, "to eat omelettes" differs from "to eat" in that it can no longer take a direct object; the reclassification of 401, accordingly, although it carries many of the indices of its 1st correlatum, will not carry 401 N1. (Note that this rule can not be generalised to say that no product can have its own correlation-number in its reclassification; for "green string" (521 N) can function as a 521 N2; "thin green string".)

Since the I_c-strings of the 1st correlatum of a certain correlation are not always the same ("to eat omelettes" will naturally have indices different

---

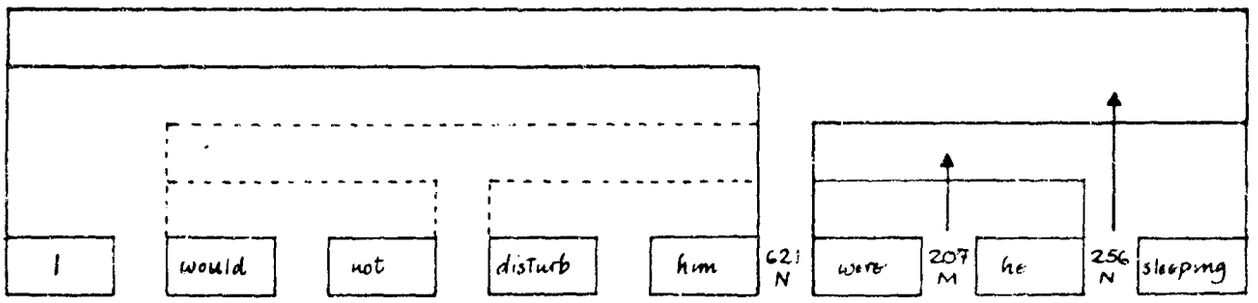
(*) The subject-verb correlations, and some correlations such as "to"//supine (=infinitive), yield entirely "new" products with possibilities quite different from those of either of their correlata. These new possibilities tend to give rise to a largely fixed list of I_c's, indifferent to variations in either correlatum.

from "eating omelettes", though the correlation is 401 N in both cases) we must choose, when some but not all of the indices of the 1st correlatum are required for reclassification of the product, between (i) enumerating the sum of the I_c's which, if present in the 1st correlatum, are to be retained, so that there are numbers of rules, each reading: "I_c n if 1st correlatum has it", and (ii) enumerating those which are not to be retained, so that the rule reads "all I_c's of the 1st correlatum except the following". The second of these looks simpler; but, in fact, which of them is more ecoromical of computer operations depends on the individual case. If only a small number of the indices of the 1st correlatum is required, it is clearly better to give a positive enumeration; if most of them are required, it is better to enumerate the exceptions.

b.) Other reclassifications similarly depend on checking whether a certain I_c is present in the strings of one correlatum, but the I_c assigned as a result of this check is not the same as the one the check looks for, which merely serves as a clue to some characteristic of the word. For example, the correlation 231 N("to"// supine) receives the reclassification 425 N. (predicative or transformativo infinitive, or present participle,// adjective) provided that its 2nd correlatum is in fact a predicative or transformative verb. But since its 2nd correlatum is a supine, it is of no use to look in its string for I_c 425 N1; the information is given us instead by the presence or absence of I_c 423 N1. So the entry in the reclassification is like this: "if 2nd correlatum bears I_c 423 N1, assign I_c 425 N1".

c) A third type of condition concerns not the I_c's of either of the correlata, but the correlation by which one of them is made. For example, products made by 256 N (all forms of "to be" except being// any present participle) receives the reclassification 621 N2 (subject + conditional // maintained subject + subjunctive) only on condition that its 1st correlatum is itself a product made by 207 M (all persons // "were" subjunctive).

This enables correlations like



to be made, whatever the I_c's of the 1st correlatum of 256 N. (Note that, in this notation, every reclassification corresponds to a movement upwards across one of the horizontal boundaries between two correlations; here, for this net to be made, 621 N2 must figure among the reclassifications of 256 N, 256 N1 must figure among those of 207 M, etc.)

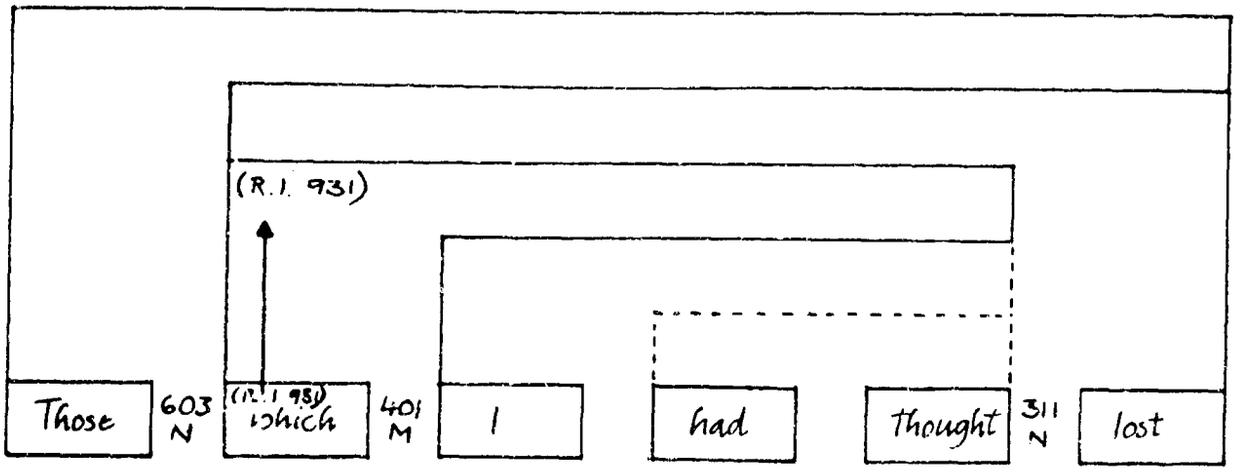
d) There are some cases in which information about a correlatum is needed for reclassification, which cannot be extracted from its own I_c's, nor by going back along the sequence of its construction even as far as the I_c's of the single words which compose it. For example, in the correlation 351 N (framing verb+

object // present participle; "I got the car//running"), only certain verbs will do; we cannot say "I made the car running", (only "I made the car run", which is a different correlation). We need, in our reclassification of 401 N, to know whether its first correlatum is one of the limited set of verbs which can make this construction. But nothing in the I_c-strings of the verb itself will tell us this, since the construction requires that a 401 N be made before the framing function becomes operative. Nor does any I_c of another number provide a criterion for this particular set of verbs. We get round this by inserting a "dummy" index (numbers between 900 and 999) at the end of the N1-string of the word "got" and of all the forms of the verbs which can make this construction. This index is not a number of a correlation, and no N2 exists to combine with it; but in the reclassification of 401 N, we can refer to it as a condition: "assign 351 N1 if 1st correlatum has R1 913".

A similar recognition index is necessary for the reclassification of relative clauses. In order to reclassify "whom I saw" as a relative clause so as to enable it to enter a correlation 603, we need to know that "whom" is a relative pronoun in the accusative - or rather, that it is one of a list of words ("whom", "which", "that") which can perform this particular function. But these words have no ordinary I_c by which they can be recognised, since their characteristic function only emerges when they have first entered a correlation 401 M. In order to be able to recognise the products 401 M which contain one of these words, we attach to "whom", "which", "that", a recognition index, R1 931. On the reclassification sheet of 401 M,

then, one entry will read like this: "if 2nd correlatum has R1 931, assign I_c 603 N2".

But it should be noted that in many cases the recognition index will be required, not at the level immediately above that on which it is originally assigned, but at some higher level, so that the recognition index has to be <u>transmitted</u>, in the reclassification of some (but not all) of the products which contain the correlatum to which it is first assigned. For example, in order to produce the analysis:



the recognition index 931, originally assigned to the word "which", must also appear among the indices of product 401 M. So, as well as the entry on the reclassification sheet of 401 M which assigns the I_c 603 N2 in virtue of the presence of R1 931, there will be another entry: "if 2nd correlatum bears R1 931, assign R1 931". And then, on the reclassification sheet of 311 N, there will be an entry: "if <u>1st</u> correlatum bears R1 931, assign I_c603 N2".

The reclassification instructions of any given correlator, then, consist of a number of indices, arranged in eight columns according to correlational functions (N1, N2, M1, M2, E1, E2, F1, F2) like the I_c-strings on word-cards. Each index except those which are to be assigned unconditionally to all products made by that correlator (the "fixed lists") has attached to it a condition determining whether it is to be assigned to the product in hand - a condition of one of the types described above.

The indices in any given string, then, can be considered in sets, where each set depends on one condition; thus, instead of reading each index, checking whether its condition is satisfied, assigning the index if the condition is satisfied, and then reading the next index, we need check each distinct condition only once, and then, if it is satisfied, assign en bloc the set of indices which depends on it.

Further, when the indices are arranged in sets in this way, we find that many sets recur, with exactly the same membership, on the reclassification sheets of more than one correlator. It is therefore practicable to set up a central store where every set of indices in use on any reclassification sheet is recorded; each entry consists of a string of indices, the condition determining their assignation, and a distinguishing code. The reclassification sheets, then, show no individual indices at all, but only, under each of the eight correlational functions, a sequence of code entries which serve as addresses to the relevant sets of indices, each set with its condition, in the central store. (See Appendix No.V) In spite of the fact that the sets in the central store are very far from mutually exclusive, this arrangement proves much more economical than the enumeration of indices on each reclassification sheet. The sets are arranged in the central store, first

in classes according to the type of condition on which they depend, indicated by a letter-code; then, within each class, according to the number of indices in the set. If, within one class, there are several sets, containing the same number of indices, they are distinguished by progressive numbering. These three levels of classification provide an unambiguous mnemonic code.

P R O G R A M O F W O R K

(Next Grant-period)

Owing to the delay of machine tests (cf. p.3-4) the most urgent task to be completed is the programming and thorough testing of the Multistore procedure. As soon as this program is running on a machine we shall be in a position to test and correct the strings of correlation indices that have been assigned to the words of the vocabulary (cf. section 5, above) as well as the system of index assignation in the reclassification procedure (cf. section 7, above).

Simultaneously the second level analysis of the explicit correlators contained in the vocabulary will be completed (cf. section 4, above). Then, while an attempt will be made to incorporate corresponding Russian output in the correlator break-up, the definitions of relations and the descriptions of correlata will be ordered with a view to establishing a preliminary skeleton of the semantic classification which is one of the targets of the project.

Subsequently the sentence analyses produced by the Multistore procedure will show in which directions this semantic classification has to be refined in order to reduce multiple output; that is, wherever the procedure yields more than one analysis for a sentence which is not ambiguous for the human reader. The reasons - provided that they spring from the sentence in hand and not from some further outside context - that enable the human reader to discard certain syntactic analyses as inappropriate will be examined in the light of the preliminary semantic classification. The classification will thus be corrected and completed to incorporate also such semantic factors as can be isolated in the field of syntactic analysis. All semantic factors,

regardless of whether they spring from correlator analysis or from the study of multiple output, will be fed back into the Multistroe system.

Finally, those cases of multiple output which the human reader resolves by reference to information gleaned, not from the sentence in hand but from the preceding text, will be studied with a view to devising a system of "semantic weights" which might help further to reduce ambiguities in the interpretation of English sentences.

REPORTS and PUBLICATIONS issued during Grant period

T 3 - "A Project for Automatic Sentence Analysis", (Ernst v. Glasersfeld), Informal Report No.1, May 1964; published (in English) in Beiträge zur Sprachkunde und Informationsverarbeitung, No.4, 1964; R. Oldenbourg Verlag. Munich (Germany).

Summary: see p. 5

T 4 - "IDAMI Language Research" (summary description of project), published in The Finite String, vol. 1, No. 6, part 1, June 1964; Western Reserve University, Cleveland (Ohio).

T 5 - Word List of the Project's Vocabulary, Appendix No.1 to Administrative Report No.1, August 1964.

Up-to-date version: see Appendix No.I of this report.

T 6 - List of Explicit and Implicit Correlators, Appendix No.2 to Administrative Report No.1, August 1964.

Up-to-date version: see Appendix No.II of this report.

T 7 - Analysis of the Correlator "about", (first draft), Appendix No.3 to Administrative Report No.1, August 1964.

T 8 - "Automatic English Sentence Analysis" (description of project), in Information Sciences 1964, AFOSR 65-0271, pp. 81-84 (Harold Wooster, Directorate of Information Sciences); Office of Aerospace Research, Washington, 1965.

T 9 - "English Prepositions in Automatic Translation" (Jehane Burns), Informal Report No.2, January 1965; to be published (in English) in Beiträge zur Sprachkunde und Informationsverarbeitung, Autumn 1965; R.Oldenbourg Verlag, Munich (Germany).

Summary: see p.15

T 10 - "MULTISTORE - A Procedure for Correlational Analysis" (Ernst v. Glasersfeld, Pier Paolo Pisani, Jehane Burns) Informal Report No.3, January 1965; published (in Italian) in Automazione e Automatismi, vol IX, No.2, April 1965; Milan (Italy).

Summary: see p. 23

A P P E N D I X No. I

The Vocabulary of the Project

(g r a m m a t i c a l l y o r d e r e d)

L I M I T E D V O C A B U L A R Y

Explicit Correlators

about	by	of	through
after	down	on	times
and	for	since	to
at	from	so	under
because	if	than	up
between	in	that	with
but	like	though	

Correlator Adverbs

about	in	to
by	on	under
down	through	up

Adverbs

away	firstly	later	pleasantly
brightly	heavily	latest	rather
darkly	here	little	so
easily	how	more	there
fast	kindly	most	very
faster	largely	newly	when
fastest	late	not	where
first	lately	now	willingly

Adjectives

bright	fastest	largest	new
brighter	first	late	newer
brightest	green	later	newest
broke	greener	latest	old
dark	greenest	like	older
darker	heavier	likelier	oldest
darkest	heaviest	likeliest	own
difficult	heavy	likely	pleasant
easier	kind	little	telling
easiest	kinder	live	white
easy	kindest	livelier	whiter
English	knowing	liveliest	whitest
fast	large	lively	willing
faster	larger		

Categorisers

a	his	some	three
an	its	that	two
every	my	the	what
few	nine	their	which
fewer	no	these	whose
fewest	one	this	your
her	our	those	

Pronouns

few	it	some	what
fewer	its	that	which
fewest	itself	theirs	who
he	me	them	whom
her	mine	these	whose
hers	myself	they	you
herself	nine	this	yours
him	none	those	yourself
himself	one	two	
his	ours	us	
I	she	we	

Auxiliaries

be	is	been	being	was
	am			were
	can			could
do	does	done	doing	did
have	has	had	having	
	may			might
	must			
	ought			
	shall			should
	will			would

Infinitive Particle

to

Verbs

answer	answers	answered		answering
ask	asks	asked		asking
book	books	booked		booking
break	breaks	broke	broken	breaking
cake	cakes	caked		caking
can	cans	canned		canning
colour	colours	coloured		colouring
down	downs	downed		downing
eat	eats	ate	eaten	eating
fast	fasts	fasted		fasting
go	goes	went	gone	going
hand	hands	handed		handing
house	houses	housed		housing
imagine	imagines	imagined		imagining
keep	keeps	kept		keeping
know	knows	knew	known	knowing
lift	lifts	lifted		lifting
like	likes	liked		liking
live	lives	lived		living
look	looks	looked		looking
make	makes	made		making
man	mans	manned		manning
mine	mines	mined		mining
own	owns	owned		owning
prepare	prepares	prepared		preparing
question	questions	questioned		questioning
rain	rains	rained		raining
read	reads	read		reading
say	says	said		saying
sit	sits	sat		sitting
tell	tells	told		telling
time	times	timed		timing
train	trains	trained		training
wait	waits	waited		waiting
walk	walks	walked		walking
want	wants	wanted		wanting
will	wills	willed		willing
wish	wishes	wished		wishing
work	works	worked		working

Nouns

(singular)	(singular Saxon g.)	(plural)	(plural Saxon g.)
Answer	answer's	answers	
being	being's	beings	beings'
book	book's	books	books'
booking		bookings	
break	break's	breaks	
cake	cake's	cakes	
can	can's	cans	
car	car's	cars	cars'
Charles	Charles's		

(nouns, continued)

colour	colour's	colours	
colouring		colourings	
dark			
day	day's	days	days'
doing		doings	
down	down's	downs	
electricity			
English		English	
fast	fast's	fasts	
find		finds	
glass		glasses	
go			
		goings	
green	green's	greens	
hand	hand's	hands	
hill	hill's	hills	
house	house's	houses	houses'
housing			
keep			
keeping			
kind	kind's	kinds	
know			
lemon	lemon's	lemons	
letter	letter's	letters	
life	life's	lives	
lift	lift's	lifts	
like		likes	
liking		likings	
living			
look	look's	looks	
make	make's	makes	
making			
man	man's	men	men's
May	May's		
might	might's		
mine	mine's	mines	
mining			
must		musts	
nine	nine's	nines	
one	one's	ones	
piece	piece's	pieces	
question	question's	questions	questions'
rain	rain's	rains	
read			
reading	readings	readings	
say			
saying		sayings	
sister	sister's	sisters	sisters'
sitting		sittings	
story	story's	stories	
table	table's	tables	tables'
tea	tea's	teas	
three	three's	threes	

(nouns, continued)

time	time's	times	
timing			
town	town's	towns	towns'
train	train's	trains	
training			
two	two's	twos	
		ups	
wait	wait's	waits	
walk	walk's	walks	
want		wants	
white	white's	whites	
will	will's	wills	
window	window's	windows	
wish		wishes	
work	work's	works	
working		workings	

APPENDIX No. II

Table of Correlators

a) Explicit Correlators contained in the Vocabulary of the Project;

b) General Table of Implicit Correlators

Explicit Correlators Contained in the Project's Vocabulary

(30 June 1965)

I_c	correlator type	correlator
003	E & F	about
007	E & F	after
014	E	and
019	E & F	at
022	E & F	because
029	E & F	between
031	E	but
032	E & F	by
036	E & F	down
040	E	for (conjunction)
041	E & F	for (preposition)
042	E & F	from
044	E & F	if
045	E & F	in
049	E & F	like
057	E & F	of
059	E & F	on
074	E & F	since (conjunction)
075	E & F	since (preposition)
076	E	so
077	E & F	than
078	E & F	that
079	E & F	though
080	E & F	through
083	E	times
084	E & F	to
086	E & F	under
092	E & F	up
098	E & F	with

BLANK PAGE

Implicit Correlators

(30 June, 1965)

SUBJECT // AUXILIARY

201 N & M	"I"	"am"	(pres.indic.)
202 N & M	"we","you","they", plural substantives	"are"	" "
203 N & M	"he","she","it", singular substantives	"is"	" "
204 N & M	all persons & subst.	"be"	(pres.subj.)
205 N & M	"I","he","she","it", singular substantives	"was"	(past indic.)
206 N & M	"we","you","they", plural substantives	"were" S1	(past indic.)
207 N & M	all persons & subst.	"were" S2	(past subj.)
208 N & M	all persons & subst.	"shall","will"	
209 N & M	all persons & subst.	"should","would","could", "might"	
212 N & M	"I","we","you","they", plural substantives	"have"	
213 N & M	"he","she","it", singular substantives	"has"	
215 N & M	all persons & subst.	"had"	
216 N & M	all persons & subst.	"can","may"	
217 N & M	all persons & subst.	"did"	
218 N & M	"he","she","it", singular subst.	"does"	
219 N & M	all persons & subst.	"must"	
220 N & M	all persons & subst.	"ought"	
221 N & M	"I","we","you","they", plural substantives	"do"	

SUBJECT // VERB

222 N	"I","we","you","they", plural substantives	supine	(pres.indic.)
223 N	"he","she" "it", singular substantives	3rd person	(pres.indic.)
225 N	all persons & subst.	past tense	

AUXILIARY COMBINATIONS

231 N	"to"		supine (infinitive)
235 N	"ought"		infinitive
		e.g.	we ought//to go
236 N	to have, all forms		infinitive
		e.g.	we have//to go
237 N	to be, all forms except "be"		infinitive
		e.g.	we are//to go
251 N	to have, all forms		"been"
		e.g.	we have//been
252 N	to be, all forms except "be","being"		"being"
		e.g.	we are//being
253 N	"shall","should","will", "would","can","could", "must","might","may"		"be"
254 N	" " " + "do","does","did"		"have"
255 N & M	" " "		supine, ordinary verbs
256 N	to be, all forms except "being"		present participle
257 N & M	to be, all forms		past participle
258 N & M	to have, all forms		past participle

FRAMING CORRELATIONS

301 N — framing verb
subj.+ indicative
e.g. I know//he liked whisky

303 N — adjectival expr. (evaluative)
subj.+ indicative
e.g. how sad//they have to go

311 N — framing verb + obj.
past part. or adj.
e.g. I consider him//beaten

315 N — framing verb + obj.
substantive
e.g. we thought him//a fool

321 N — framing verb
subject + subjunctive
e.g. I wish//he would go

331 N — framing ~~verb~~ noun
infinitive
e.g. there is no need//to leave

333 N — framing adj. or past part.
infinitive
e.g. John was hard//to convince
(Note: the adj. modifies the inf.)

335 N — adjectival expr. (evaluative)
infinitive
e.g. it is difficult//to convince him
(Note: the adj. is predicative)

337 N — "time"
subject + subjunctive
e.g. it is time//we went

[338 N] — "the way"
subject + indicative
e.g. the way//he does it

339 N — "the moment","the day","the time"
subject + indicative
e.g. the day//we landed

343 N — framing verb
infinitive
e.g. they started//to run

345 N — framing verb
present participle
e.g. he started//crying
(non-contemporary activities)

351 N — framing verb + obj.
present participle
e.g. I got the car//running
(Note: p.p. indicates activity of the object; similar to "I got the car to run")

353 N framing verb + obj. past participle

e.g. I got the car//cleaned

(Note: the object is object also of the activity indicated by the p.p.)

355 N framing verb + obj. infinitive

e.g. we got the doctor// to come

(Note: the p.p. indicates activity of the object)

357 N framing verb + obj. supine

e.g. we saw the guests//leave

SPECIFICATION BY A VERB

361 N substantive infinitive + obj.(purpose)

e.g. a stick//to beat the dog

363 N to be, personal form infinitive (purpose)

e.g. these apples are//to eat

365 N substantive infinitive (purpose)

e.g. milk//to drink

367 N subject + indic. infinitive (purpose)

e.g. he works//to live

373 N & M subject + verb past partic. or adj.

e.g. he lay//exhausted

381 N subject + fr. verb supine

e.g. the animal he let//run

382 N subject + fr.verb infinitive

e.g. the bear he taught//to dance

383 N subj.+ fr. or transf. verb past part. or adj.

e.g. the man they considered//beaten

384 N subj.+ fr. or transf. verb present participle

e.g. the car he got//running

385 N subj. + transf. verb noun

e.g. the man they made//king

391 N transformative verb + object substantive or adjective

e.g. they made him//president

395 N subject + verb present participle

e.g. we arrived//singing

VERB // OBJECT

401 N & M	verb		object (accusative)
		e.g.	John hates//her
411 N	verb		object (dative)
		e.g.	we wrote//you a card

VERB // PREDICATE

421 N & M	predicative verb		substantive
		e.g.	he was//president
422 N & M	to have, all forms		substantive
		e.g.	they have//illusions
423 N & M	predicative or transf. verb, personal forms		adjective
		e.g.	we were//cold
425 N	pred. or transf.inf. or present part.		adjective
		e.g.	to be//cold
429 N	pred. or transf.p.p.		adjective
		e.g.	it must have seemed//strange

MODALISING CORRELATIONS

441 N & M	verb		adverb
442 N & M	verb		"here","there","where"
443 N & M	verb		correlator adverb
		e.g.	they came//in
445 N & M	adverb		adjective
447 N	modifying adv.		adverb
449 N	substantive		correlator adverb
		e.g.	the way//out
451 N	quantitative subst.		substantive (unit)
		e.g.	3 miles//an hour
455 N	adverb		subject + verb
		e.g.	suddenly//they left
457 N	relative adverb		noun, adj., pres. part. or p.p.
		e.g.	when//drunk John is a bore

461 N	adverb		demonstrative adverb
		e.g.	we are nearly//there
462 N	correlator adverb		demonstrative adverb
		e.g.	we found them in//there
463 N	adverb		correlator
		e.g.	he lives almost//in Paris
465 N	adverb		correlator adverb
		e.g.	he was nearly//down
467 N	adverb		comparative adjective
469 N	adverb		comparative adverb
471 N	adverb		infinitive
		e.g.	nearly//to win is not enough
473 N & M	adverb		substantive
		e.g.	**he thought her quite//a girl**
481 N	auxiliary		negation
		e.g.	you are//not a baby
483 N	negation		pres. or past participle
		e.g.	not//having succeeded, Charles left
485 N	negation		infinitive
		e.g.	it's unforgivable not//to try
487 N	negation		supine
		e.g.	she should walk and not//run

NOUN CORRELATIONS

501 N	definite article		noun
503 N	indefinite article		noun
505 N	categorising adject.		noun
		e.g.	these//people; our//car
513 N	"you" or "we"		noun or nominalised adj.
		e.g.	you//English; we//poor
515 N	Saxon genitive		noun
		e.g.	John's//car
521 N	adjective		noun
		e.g.	her green//hat

[522 N] indefinite pronoun — adjective

e.g. something//ugly

523 N substantive — adj., pres. part. or past participle

e.g. we bought the house//unfinished

(Note: the p.p. modifies the substantive, not the preceding subject; cf. 373)

[526 N] substantive -,- — adjectival expression

e.g. the guests/,/noisy and full of zest...

528 N comparative + "the" — substantive

e.g. the later//the drawings, the cheaper they are

536 N substantive — reflexive pronoun

e.g. the man//himself

544 N substantive — "on" or "off"

e.g. he came in with his hat//on

PROXIMITY CORRELATIONS

551 N noun — noun

e.g. danger//level

553 N noun — past participle

e.g. home//baked

555 N noun — adjective

e.g. ice//cold

557 N adjective — adjective

e.g. red//hot

559 N supine — noun

e.g. wash//room

561 N present participle — noun

e.g. dining//table

563 N present participle) — adjective

e.g. fighting//mad

COINCIDENCE CORRELATIONS (Apposition)

[571 N]	substantive	-,-	common noun
			e.g. the soup/,/a greasy mess
			(Note: the correlation establishes an identity)
[573 N]	specific term or proper name	-,-	specific term or proper name
			e.g. the president/,/Mr Smith
			(Note: the correlation expressees a pre-established identity)
575 N	specific term or proper name		specific term or proper name
			e.g. the president//John Smith
577 N	title		proper name
			e.g. Doctor Finnegan
579 N	proper name		proper name
			e.g. John//Smith

RELATIVE CLAUSES

601 N	substantive	subject + transitive, framing, or transf. verb
		e.g. the man//we saw;
		the chair//he painted red;
602 N	substantive (subject of the clause)	subj. + pred.verb
		e.g. the man//he was
603 N	substantive (obj. of the clause)	rel. pron.+ subj. + trans.verb
		e.g. the man//whom we saw
605 N	substantive (subj. of the clause)	rel. pron.+ verb
		e.g. the car//that crashed
607 N	substantive (specified by the clause)	rel.adverb+subj.+verb
		e.g. the place//where he lived
[608 N]	subject + verb -,-	"which" + predicative, transformative, or transitive verb and complement
		e.g. he came to Rome/,/which was nice

609 N	substantive	1st correlatum of explicit correlator + correlator
		e.g. the house//we live in
611 N	substantive	rel.pron.+ 1st correlatum of explicit correlator + correlator
		e.g. the house//which we live in
613 N	substantive	explicit correlator + rel.pron. + 1st correlatum of explicit correlator (F-correlation)
		e.g. the house//in which we live

CONDITIONAL CORRELATIONS

621 N & M	subject + conditional	subject + past subjunctive
		e.g. I should eat//were I hungry

ABLATIVE ABSOLUTE

631 N*	subject + pres.participle	subject + indicative
		e.g. John being asleep//we left
633 N & M	pres.participle	subject + indicative
		e.g. running down the hill//he _ _ broke his leg
635 N*	past part.or adj.	subject + indicative
		e.g. tanned//he looked better

* Maintainment (inversion) requires a comma.

Note: square brackets indicate that the correlation requires either a comma or a word that is not contained in our vocabulary.

A P P E N D I X No. III

C o r r e l a t o r A n a l y s i s

Sample of second-stage analysis:

Break-up of the Correlator "BY"

1st correlatum	2nd correlatum	CORRELATOR
substantive	substantive, fixed point of spatial reference, spatially limited, vertical.	BY I_c 032
relation		
spatial proximity		I_r 011

e.g.

a chair by the window

the house by the church

Note: this relation of particular closeness seems to require one of two conditions:

coincidence of one vertical boundary,

or

absence of vision-obstructing obstacle between the two correlata.

Italian OUTPUT

accanto a

German

bei

French

près de

1st correlatum	2nd correlatum	CORRELATOR
activity	substantive, fixed point of spatial reference, spatially limited, vertical.	BY I_c 032
relation		
spatial proximity		I_r 012

e.g.

to sit by the fire

he stood by the big tree

Note: this relation seems to require:

absence of vision-obstructing obstacle between the two correlata.

Italian OUTPUT

accanto a

German

bei

French

près de

1st correlatum	2nd correlatum	CORRELATOR
substantive	substantive, point of spatial reference, unlimited at least in one dimension.	BY I_c 032
relation spatial proximity		I_r 021
e.g. the house by the river a village by the sea		**Italian** OUTPUT vicino a **German** an **French** auprès de

1st correlatum	2nd correlatum	CORRELATOR
activity	substantive, point of spatial reference, unlimited in at least one dimension.	BY I_c 032
relation spatial proximity		I_r 022
e.g. she fell by the wayside we live by the canal		**Italian** OUTPUT vicino a **German** an **French** auprès de

1st correlatum	2nd correlatum	CORRELATOR
locomotive activity (of subject or object)	substantive, dat. pron., point of spatial ref. spatially limited.	BY I_c 032
relation		
momentary spatial proximity		I_r 031

e.g.

we drove by a large church

he walked by me without smiling

Note: in Italian this construction is possible only with the verb "passare" and "vicino a" or "accanto a". The same goes for French.

OUTPUT	
Italian	transf.
German	an .... vorbei
French	transf.

1st correlatum	2nd correlatum	CORRELATOR
locomotive activity (of subject or object)	substantive, geogr. name, point of spatial ref. spatially limited.	BY I_c 032
relation		
temporary spatial coincidence, determining itinerary (="via")		I_r 041

e.g.

the guests came by the village

he went to Rome by Bologna

Note: the locomotive activity must either imply (to come, to arrive, to leave, to depart, etc.) or be specified by, the starting point or the end of the

OUTPUT	
Italian	transf
German	über
French	par

1st correlatum	2nd correlatum	CORRELATOR: BY
substantive implying locomotion	substantive, geogr. name, point of spatial reference, spatially limited.	I_c 032
relation: temporary spatial coincidence determining itinerary (="via")		I_r 042

e.g.

their arrival by the fields

the journey by Panama is longer

OUTPUT

Italian: per

German: über

French: par

1st correlatum	2nd correlatum	CORRELATOR: BY
activity, terminative	substantive, temporal specification (limit).	I_c 032
relation: temporal limitation		I_r 051

e.g.

get it ready by Friday

we should finish this by July

OUTPUT

Italian: entro

German: bis

French: jusqu'à

1st correlatum	2nd correlatum	CORRELATOR: BY
activity, terminative.	name of season, (spring, summer, fall autumn, winter).	I_c 032
relation: temporal limitation		I_r 052
e.g.: get it ready by the summer; he had returned by the winter		**OUTPUT** — Italian: entro; German: bis zu; French: jusqu'à

1st correlatum	2nd correlatum	CORRELATOR: BY
activity, non-terminative	substantive, temporal specification, (limit).	I_c 032
relation: temporal limitation		I_r 053
e.g.: I was tired by the evening; we were asleep by midnight		**OUTPUT** — Italian: ore di; German: transf.; French: transf.

1st correlatum	2nd correlatum
activity, implying conventional units of some kind.	substantive, unit of space, time, number, or price.

CORRELATOR

BY

I_c 032

relation

modalisation of activity

I_r 061

e.g.

to hire a boat by the hour

the price was calculated by the gallon

OUTPUT

Italian

a

German

pro

French

à

1st correlatum	2nd correlatum
activity, not implying units.	substantive, unit of space, time, number, or price.

CORRELATOR

BY

I_c 032

relation

modelisation of activity

I_r 062

e.g.

he drinks whisky by the gallon

we wasted tape by the yard

*"literweise", "meterweise", etc.

Note: the units given by the second correlatum are applied directly to the object of the activity and do not refer to a conventional correspondence as in the

OUTPUT

Italian

transf.

German

adverb*

French

transf.

1st correlatum	2nd correlatum	CORRELATOR
activity, passive construction	substantive, conative subject of the activity	BY I_c 032
relation		
activity - its agent		I_r 071
e.g.		**OUTPUT**
the window was opened by the maid		**Italian** da
the steak was eaten by a dog		**German** von
he was despised by women		**French** par

1st correlatum	2nd correlatum	CORRELATOR
activity, passive construction	substantive, unintentioned subject of the activity	BY I_c 032
relation		
activity - its subject		I_r 072
e.g.		**OUTPUT**
the window was opened by the wind		**Italian** da
we shall be ruined by the slump		**German** durch
		French par

1st correlatum	2nd correlatum	CORRELATOR: BY
activity, passive construction	substantive, instrument of the activity, "means" implying intention	I_c 032
relation activity - its means		I_r 081
e.g. the boat was propelled by oars he was killed by an arrow the shelf is held by a screw		**OUTPUT** **Italian** da **German** durch **French** par

1st correlatum	2nd correlatum	CORRELATOR: BY
substantive, nomen actionis	substantive, instrument of the activity, "means" implying intention	I_c 032
relation activity - its means		I_r 082
e.g. propulsion by oars death by an arrow		**OUTPUT** **Italian** mediante **German** durch **French** par

1st correlatum	2nd correlatum	CORRELATOR: BY
activity, passive construction	substantive, instrument subject of the activity, unintentioned.	I_c 032
relation activity - its means		I_r 083

e.g.

he was killed by an arrow

it was shifted by a **lever**

Note: in German the construction with "von" implies, unlike that with "durch" (see 082), that no intention is involved.
In English this can be rendered only explicitly, by an adverb like "accidentally".

OUTPUT	
Italian	da
German	von
French	par

1st correlatum	2nd correlatum	CORRELATOR: BY
substantive, product, artifact	proper name, producer, author	I_c 032
relation authorship		I_r 091

e.g.

a play by Shakespeare

a picture by Gauguin

a garage by Frank Lloyd Wright

OUTPUT	
Italian	di
German	von
French	de

1st correlatum	2nd correlatum	CORRELATOR
activity, transformative	nomen actionis, method employed (intentioned)	BY I_c 032
relation result - method		I_r 101
e.g. the village was destroyed by bombing he was cured by hypnosis Note: distinguishable from 072 and 102 by the possibility of introducing a subject: e.g. "They destroyed the village by bombing"		**Italian** OUTPUT con **German** durch **French** par

1st correlatum	2nd correlatum	CORRELATOR
activity, passive construction	nomen actionis, unintentioned activity of the subject of the 1st c.	BY I_c 032
relation result - cause		I_r 102
e.g. he was betrayed by his stammer		**Italian** OUTPUT da **German** durch **French** par

1st correlatum	2nd correlatum	CORRELATOR
activity (mental?)	substantive implying specific mental activity or method (intentioned)	BY I_c 032
relation result - method		I_r 103
e.g. he knew by intuition to choose by elimination it follows by deduction one can prove it by analogy		**Italian** OUTPUT per **German** durch **French** par

1st correlatum	2nd correlatum	CORRELATOR
nomen actionis (mental)	substantive, implying specific mental activity or method	BY I_c 032
relation result - method		I_r 104
e.g. proof by analogy choice by elimination		**Italian** OUTPUT per **German** durch **French** par

1st correlatum	2nd correlatum	CORRELATOR: BY
locomotive activity of subject	substantive, means of transport, with fixed itinerary	I_c 032
relation: locomotion - means		I_r 110
e.g.: they arrived by train		**Italien** OUTPUT: in **German**: per **French**: par

1st correlatum	2nd correlatum	CORRELATOR: BY
locomotive activity of subject	substantive, means of transport without fixed itinerary	I_c 032
relation: locomotion - means		I_r 111
e.g.: to travel by car		**Italien** OUTPUT: in **German**: per **French**: en

1st correlatum	2nd correlatum	CORRELATOR: BY
locomotive activity of object	substantive, means of transport	I_c 032
relation: locomotion - means		I_r 112
e.g. send the letter by air mail		**Italian** OUTPUT: per **German**: per **French**: par

1st correlatum	2nd correlatum	CORRELATOR: BY
activity, entailing physical or mental contact	substantive, instrumental part or characteristic of object, (not agent)	I_c 032
relation: specification of contact		I_r 121
e.g. he is to be hanged by the neck catch a cat by its tail I recognised him by his stammer* *Note: different from "betrayed by his stammer", because the stammer cannot be the active subject of the activity given by the 1st c.		**Italian** OUTPUT: per **German**: an **French**: par

1st correlatum	2nd correlatum	CORRELATOR
person or personal aspect	substantive, dispositional cause appertaining to 1st c	BY I_c 032
relation aspect - its cause (rather like "owing to")		I_r 131

e.g.

a democrat by conviction

a victim by preference

a philosopher by inclination

Italian	OUTPUT
per	
German	
aus	
French	
par	

1st correlatum	2nd correlatum	CORRELATOR
person or personal aspect	substantive, efficient cause outside 1st c.	BY I_c 032
relation aspect - its cause		I_r 132

e.g.

he was made consul by appointment

she became Italian by marriage

a leader by general acclamation

Italian	OUTPUT
per	
German	
durch	
French	
par	

1st correlatum	2nd correlatum	CORRELATOR
activity	specific substantive "day", "night"	BY I_c 032
relation temporal condition		I_r 141
e.g. we travelled by night they attacked by day		**Italian** OUTPUT di **German** bei **French** pendant

1st correlatum	2nd correlatum	CORRELATOR
substantive	specific substantive "day", "night"	BY I_c 032
relation temporal condition		I_r 142
e.g. a journey by night an attack by day		**Italian** OUTPUT di **German** bei **French** pendant

1st correlatum	2nd correlatum
activity	substantive specification of illumination

CORRELATOR

BY

I_c 032

relation

condition of illumination

I_r 151

e.g.

he reads by candle light

they danced by moonlight

Italian OUTPUT

a

German

bei

French

à

1st correlatum	2nd correlatum
substantive	substentive, specification of illumination

CORRELATOR

BY

I_c 032

relation

condition of illumination

I_r 152

e.g.

a party by candle light

Italian OUTPUT

a

German

bei

French

à

1st correlatum	2nd correlatum	CORRELATOR
number	number	BY
		I_c 032
relation: multiplication		I_r 160
e.g. two by three		Italian: per German: mal French: fois OUTPUT

1st correlatum	2nd correlatum	CORRELATOR
dimension, or measurable concept	dimension, or measurable concept	BY
		I_c 032
relation: multiplication		I_r 161
e.g. length by width		Italian: per German: mal French: per OUTPUT

1st correlatum	2nd correlatum
specific activity "to multiply"	number, dimension, or measurable concept

CORRELATOR

BY

I_c 032

relation

multiplication

I_r 162

e.g.

to multiply two by three

to multiply pressure by surface

Italian OUTPUT

per

German

mit

French

par

1st correlatum	2nd correlatum
specific activity "to divide"	number, dimension, or measurable concept

CORRELATOR

BY

I_c 032

relation

division

I_r 171

e.g.

to divide four by two

to divide width by length

Italian OUTPUT

per

German

durch

French

par

1st correlatum	2nd correlatum	CORRELATOR
statement	substantive, source of information	BY
		I_c 032
relation		
specification of source		I_r 181

e.g.	Italian	OUTPUT
it's six o'clock by my watch		secondo
by his account she arrived yesterday	German	
		nach
	French	
		selon

A P P E N D I X No. IV

Sample of Word-cards

(I_c-Assignation)

Correlation indices of the word "HE":

The indices are divided into 8 strings according to the correlator type.

N1		N2	M1		M2
203	218	421	203	218	421
204	219	473	204	219	473
205	220		205	220	
207	223		207		
208	225		208		
209	536		209		
213	603		213		
215	605		215		
216	609		216		
217	611		217		
	613				

E1		E2	F1	F2
003	045	014	---	---
007	049	031		
014	057	044		
019	059	077		
029	075			
031	079			
032	080			
036	084			
041	086			
044	092			
	098			

Correlation indices of the word "WERE" S1:

N1	N2	M1	M2
237	206	441	206
252		442	
256			
257			
421			
423			
441			
442			
443			
481			

E1		E2	F1	F2
003	045	014	---	---
007	049	031		
014	057			
019	059			
022	075			
029	079			
031	080			
032	084			
036	086			
041	092			
044	098			

Correlation indices of the word "WERE" S2:

N1	N2	M1	M2
237	207	441	207
252		442	
256			
257			
421			
423			
441			
442			
443			
481			

E1		E2	F1	F2
003	045	014	---	---
007	049	031		
014	057			
019	059			
022	075			
029	079			
031	080			
032	084			
036	086			
041	092			
044	098			

Correlation indices of the word "SLEEPING":

N1	N2	M1	M2
441	256	441	---
442	345	442	
443	351		
	395		
	457		
	523		

E1	E2	F1	F2
003 045	003 057	003 057	003 075
007 049	007 059	007 059	007 077
014 057	014 075	019 074	019 079
019 059	019 077	022 075	022 080
022 074	022 079	029 079	032 084
029 075	031 080	032 080	041 098
031 079	032 084	036 084	044
032 080	041 098	041 086	045
036 084	044	044 092	049
041 086	045	045 098	057
044 092	049	049	059
098			

APPENDIX No. V

Sample of Reclassification

Reclassification Sheet for Products of

CORRELATION 207 M:

2	0	7	M

N 1		N 2		M 1		M 2	
	FL 11/1		FL 1/5		FL 3/6		FL 1/5
E 1		**E 2**		**F 1**		**F 2**	
	FL 14/2		---		---		---

The code numbers on this sheet address the following reclassification lists:

FL 1/5: 621

FL 3/6: 421, 423, 455

FL 11/1: 237, 252, 256, 257, 363, 421, 423, 442, 443, 455, 481

FL 14/2: 003, 007, 029, 032, 036, 042, 045, 049, 057, 059, 080, 086, 092, 098

All of them are "Fixed Lists", i.e. the assignation of the listed I_c's is unconditional and any product of the correlator 207 M receives these I_c's.

Reclassification Sheet for Products of

CORRELATION 256 N:

2 | 5 | 6 | N

N 1		N 2		M 1		M 2	
	RA 1/10		LA 9/1 RA 1/9 RA 11/1		FL 1/7		RA 1/9
E 1		**E 2**		**F 1**		**F 2**	
	FL 2/1		FL 2/1 RA 1/3		---		---

The code numbers on this sheet address the following reclassification lists:

FL 1/7: 441

FL 2/1: 014, 031

These are "Fixed Lists", i.e. lists of I_c's which are assigned unconditionally to any product of the correlator 256 N.

LA 9/1: 201, 202, 203, 204, 205, 206, 207, 251, 253,

The I_c's contained in lists LA are assigned to the product, if they are part of the corresponding string of the product's first correlatum.

RA 1/3: 077

The I_c's contained in lists RA are assigned to the product only if a given condition is satisfied. In the case of list RA 1/3 the condition is that the product was produced by correlator 231 N.

RA 1/9: 621

Assignation on condition that the product was produced by correlator 207 M.

RA 1/10: 952 (recognition index)

Assignation on condition that the product was produced by correlator 207 M.

RA 11/1: 235, 236, 237, 331, 333, 335, 343, 355, 361, 471, 485

Assignation on condition that the product's first correlatum has index 951 in string N1.

GENERAL BIBLIOGRAPHY

Note: in this bibliography we have listed the books, reports, and papers that we found most interesting and stimulating with regard to the particular field of our research.

GENERAL BIBLIOGRAPHY

Agricola, Erhard, "Automatische Sprachübersetzung, Englisch-Deutsch", Technical Report, Arbeitsstelle für Mathematische und Angewandte Linguistik, Deutsche Akademie der Wissenschaften, Berlin, 1963.

Alexander, D., and Matthews, P.H., Adjectives before That-Clauses in English", Linguistic Research Project, Indiana University, 1964.

Alexander, D., and Kunz, W.J., "Some Classes of Verbs in English", Linguistic Research Project, Indiana University, 1964.

Benzécri, J.P., et al., "Imago Primi Anni", Faculté des Sciences de l'Université der Rennes, 1964.

Bourquin, Guy, "Le groupement nominal en anglais écrit", Faculté des Lettres et Sciences Humaines, Université de Nancy, 1964.

Brown, A.F., "English Canonical Forms and Style Levels", Center for the Information Sciences, Lehigh University, Bethlehem, Pennsylvania, 1964.

Ceccato, Silvio, et al., "Linguistic Analysis and Programming for Mechanical Translation", Technical Report, RADC-TP-60-18, Centro di Cibernetica e di Attività Linguistiche, Milan University (Italy),1960.

- "Mechanical Translation: the Correlational Solution", Technical Report, RADC-TR-63-xx, ibid.., 1963.

Ceccato, Silvio, "Les limites actuelles de la traduction automatique - comment et pourquoi ne réussit-on pas à traduire?", A.I.D. Bulletin, III, 1964.

Cherry, Colin, "On Human Communication", Science Editions, Inc., New York, 1961.

Coyaud, Maurice, "Document Automatic Indexing with the Help of Semantic Information", Section d'Automatique Documentaire, CNRS, Paris (France), 1965.

Gardin, J.C., "SYNTOL - Syntagmatic Organization Language", paper prepared for the Rutgers Seminars on Systems for the Intellectual Organization of Information, Rutgers, The State University, New Brunswick, New Jersey, 1963.

Garvin, Paul L., "The Georgetown-IBM Experiment of 1954: An Evaluation in Retrospect", paper presented at the AMTCL Meeting, Bloomington, Indiana, 1964.

- (editor), "Natural Language and the Computer", McGraw-Hill Book Company, Inc., New York, 1963.

- "Inductive Methods in Semantic Analysis", Technical Notes Nos. 1-8, Thompson Ramo Wooldridge, Inc., Computer Division, Canoga Park, Calif., 1962-64.

Garvin, Paul L., et al."Fulcrum Techniques in Semanic Analysis", Final Report,RADC-TDR-63-168, Thompson Ramo Wooldridge, Inc., TRW Computer Division, Canoga Park, Calif., 1963.

Ginsburg, Seymour, and Spanier, Edwin H.,"Mappings of Languages by Two-Tape Devices", Scientific Report No.9, TM-738/007/00, System Development Corporation, Santa Monica, Calif., 1964.

Goldwin, A.J.,"The Place of Indexing in the Design of Information Systems Tests", Technical Report No.3, CSL:TR-3, Center for Documentation and Communication Research, Western Reserve University, Cleveland, Ohio, 1964.

Harper, Kenneth, E.,"Prepositional Phrases and Automatic Parsing",Memorandum-4383-PR, The Rand Corporation, Santa Monica, California, 1965.

Harris, Zellig S.,"Strings and Transformations in Language Description", University of Pennsylvania, Department of Linguistics, Philadelphia, 1961.

Harwood, F.W.,"On the Design of a Formal System for English", The University of Tasmania, May, 1964.

Hays, David G.,"Report of a Summer Seminar on Computational Linguistics", Memorandum- RM-3889-NSF, The Rand Corporation, Santa Monica, Calif.,1964.

- "Dependency Theory: a Formalism and Some Observations", Memorandum RM-4087-PR, The Rand Corporation, Santa Monica, California, 1964.

- "An Annotated Bibliography of Publications on Dependency Theory", Memorandum RM-4479-PR, The Rand Corporation, Santa Monica, Calif., 1965.

Hoppe, Alfred,"Der sprachliche Formulierungsprozess in den Funktionsebenen der Sprache", Beiträge zur Sprachkunde und Informationsverarbeitung, No.4, 1964.

- "Der sprachliche Formulierungsprozess als Grundlage automatischer Hin- und Her-Uebersetzung", Neuere Ergebnisse der Kybernetik, R.Oldenbourg Verlag, München (Germany), 1964.

IBM Watson Research Center,"Word Analyzer", Final Report RADC-TDR-62-105, April 1962.

Jenkins, R.A.,"On Conjunctive Nominalizations in English", Research Report, Air Force Cambridge Research Laboratories, OAR, Boston, Mass., 1962.

Kay, Martin, "A Parsing Program for Categorial Grammars", Memorandum RM-4263-PR, The Rand Corporation, Santa Monica California, 1964.

Katter, Robert V., "Language Structure and Interpersonal Commonality", Journal of Verbal Learning, vol.3, No.2, April 1964.

Katz, Jerrold J., and Fodor, Jerry A.,"The Structure of a Semantic Theory", Language, 39, 1963.

Katz, Jerrold J., and Postal, Paul M.,"An Integrated Theory of Linguistic Description", R.M. No.26, The M.I.T. Press, Cambridge, Mass., 1964.

Kirsch, Russel A., "Computer Interpretation of English Text and Picture Patterns",National Bureau of Standards Report, April, 1964.

Kugel, Peter,"Some Remarks on the Structures of Semantic Theories", paper presented at AMTCL Meeting, Bloomington, Indiana, 1964.

Kuno, Susumu, and Oettinger, Anthony G.,"Syntactic Structure and Ambiguity of English", Proceedings, Fall Joint Computer Conference, 1963.

Lamb, Sydney M.,"The Sememic Approach to Structural Semantics", American Anthropologist, vol.66, No.3, part 2, June, 1964.

Linguistics Research Center, University of Texas, Technical Reports, P 16, SR 1, P 22, NSF-23,(1963-65).

Masterman, Margaret M.,"Research on Semantic Basis of Communication",Annual Summary Report,Cambridge Language Research Unit, Cambridge (GB), April, 1964.

- "Semantic Basis of Communication", Final Report, Cambridge Language Research Unit, Cambridge (GB), December, 1964.

McIntosh, Angus,"Patterns and Ranges", Language, vol.37, No.3, 1961.

Melton, Jessica S.,"A Use for the Techniques of Structural Linguistics in Documentation Research", Technical Report No.4, Center for Documentation and Communication Research, Western Reserve University, Cleveland, Ohio, 1964.

Olney, John C.,"Some Patterns Observed in the Contextual Specialization of Word Senses", Information Storage and Retrieval, vol.2, Pergamon Press, Oxford (GB), 1964.

Prince, M.S., and Ryan, K.W.,"Description of English Syntactic Classes",Linguistics Research Center, The University of Texas, July, 1964.

Quillian, Ross,"A Notation for Representing Conceptual Information: an Application to Semantics and Mechanical English Paraphrasing", SP-1395, System Development Corporation, Santa Monica, Cal., 1963.

Ramo Wooldridge,"Machine Translation Studies of Semantic Techniques", Final Report, RADC-TR-61-72, Thompson Ramo Wooldridge Inc., Canoga Park, California, 1961.

Robinson, Jane," Endocentric Constructions and the Cocke Parsing Logic", paper presented at the International Conference on Linguistics, New York, 1965.

Ross, Douglas T., "On Context and Ambiguity in Parsing", ACM, vol.7, No.2, February, 1964.

Saporta, Sol, (editor),"Psycholinguistics", Holt Rinehart and Winston, New York, 1961.

Schwartz, Eugene S., "English Word Count" (Appendix A - Alphabetical Listing of Experimental Dictionary), I.I.T., Research Institute, 1963.

Sebeok, Thomas A., "Information Structures in Linguistics", Final Report, AFORS 72-63, Research Center in Anthropology, Folklore and Linguistics, Indiana University, Bllomington, Indiana, 1963.

Smith, Carlotta S., "A Class of Complex Modifiers in English", Language, vol.37, No.3, 1961.

Tabory, Robert, "Semantics, Generative Grammars and Computers", Research Report RC-1108, IBM Watson Research Center, New York, 1963.

Terzi, Paolo, "Outline for the Experimental Testing of Scientific Classifications", paper presented at the International Study Conference on Classification Research, Elsinore, September, 1964.

Wängler, Hans-Heinrich,"Rangwörterbuch hochdeutscher Umgangssprache", N.G. Elwert Verlag, Marburg, 1963.

Wilks, Shillan, Dobson, and Sparck Jones, "Semantic Basis of Communication", Final Report, Cambridge Language Research Unit, Cambridge (England), 1964.

Wooster, Harold, "A Web of Words", AFOSR, Directorate of Information Sciences,Washington, 1963.

Yngve, V.H., "A Framework for Syntactic Translation", Mechanical Translation, IV, No.3, M.I.T., Cambridge, Mass., 1957.

- "A Model and an hypothesis for Language Structure", Technical Report No.369, M.I.T., Cambridge Mass., August, 1960.

- "Random Generation of English Sentences", Proceedings of the 1961 International Conference on Machine Translation of Languages and Applied Language Analysis, N. P. L., Teddington, England, 1962.

PART II: Research of the MATHEMATICS GROUP

Principal Investigator: Paolo Terzi

Mathematician: Franco Rusconi

Programming expert: Mario Baroggi

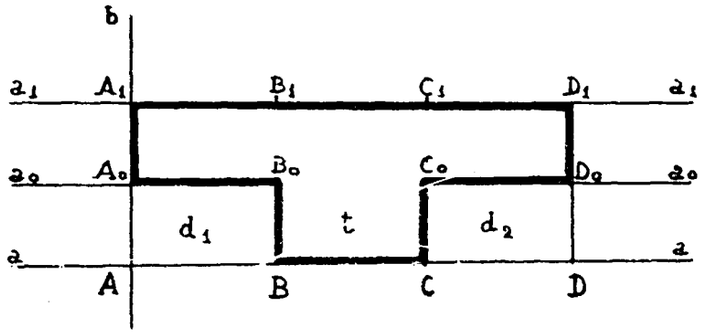
BLANK PAGE

INTRODUCTION

Our research into the mathematical aspects of the problem of representing language in such a way that parts of the meaning of a discourse can be made accessible to a machine was discussed at the Congress of Elsinore. This research has occupied all our mathematical researchers up to the present. Our studies have concerned geometrical representation of linguistic structure on the one hand, and algebraic and nomographic notation on the other.

Had any one of the specific problems which arose been immediately pursued to the end, there would have been a danger of an exaggeratedly one-sided approach. Instead we have tried initially to arrive at an overall picture of the various possible facets of the research in hand, by carrying out a series of pilot-studies which can subsequently be carried further.

The basis of our research has been the attempt to apply the basic structural unit, which we represent thus :



in such a way as to yield the maximum information for the purposes of natural language analysis and processing.

We have explored numerous ways of applying this structure, by mathematical or by algebraic means, and for 16 of these ways a preliminary investigation has been carried out which permits us to present 16 distinct studies. These 16 partial investigations will be enumerated later.

FUNDAMENTAL RESEARCH ALREADY CARRIED OUT.

All our enquiries, actual or possible, must employ a common structure. The attempt to set up a basic architectonics of language has been the object of a separate line of research which seeks to determine the structure and the origins of both thought and language.

It was necessary to find a basis common to the individual elements of the geometrical and algebraic research on the one hand; and the linguistic items themselves on the other.

A wide enquiry was made, with this problem in mind, in that section of philosophical literature which deals with the origins of ideas and of mental connections. This enquiry is reported in a paper of several chapters, at present being completed; a résumé is appended to this Report. Parts of this discussion are to be published in Italian, and an English version is in preparation, which will be published under grant 65-76.

The study of the origin of ideas as a psychological problem may seem a somewhat gratuitous preamble to the study of language, particularly the mathematical study of language; but in fact this is not the case.

Our mathematical study of language is based on the geometrical representation of correlational "nets". But to give a full definition of such a "net" we must know the nature and not only the pattern of the connections between the mental items deployed in it - from the simplest items to the most complex - and, accordingly, we must know to what connections a given mental item may, of its own nature, be subject.

Every "idea" carries the stamp of its origin. We must have a clear conception, then, of how ideas are formed. The three basic sources of ideas :

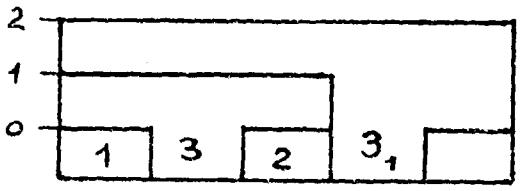
- sense-impressions;
- reflection;
- inter-personal communication;

must provide us with the data to establish the hierarchies and the values of the mental connections, with certainty and relative precision; for if this certainty and precision were not inherent in the data, language could not function as a means of communication.

The hierarchic arrangement of mental connections, and their individual values, provide in fact the content of a geometric scheme which is designed to go beyond the simple graphic representation of words, to give a representation of those connections in thought which enable a string of mental items to become synthesised in a complete "idea".

1) - GEOMETRICAL CHARACTERISTICS OF CORRELATIONS.

In the geometrical notation:



we take into consideration the following factors :

1) length + thickness of the arms of the T-shaped area $3, 3_1$, which represents the relation ;

2) area subtended beneath each arm, and total area of the rectangle;

3) maximum + minimum total length of each arm, and of the correlation rectangle as a whole;

4) ratio between the "degree", (i. e. the height of the T. The unit is the height of the designatum, and also of each of the levels above it. Thus the height, or degree, of the correlation 3 above, = 1; that of 3_1 =2.) of the correlation, the number of designata or items correlated; (small rectangles 1 and 2) and the thicknesses of the arms.

This study has been carried far enough to provide a structure in which

certain fundamental relationships can be displayed:

- space-action;
- time-action;
- space-subject;
- time-subject;
- space-development;
- time-development;
- etc.

In a further enquiry along these lines we should seek to establish whether the locations of these elements can in practice be exploited for purposes of meaning - recognition. No programs to this effect have yet been carried out.

2) MATHEMATICAL ANALYSIS OF LANGUAGE

This study represents a first practical attempt to introduce elements of algebraic calculus into linguistic analysis. We illustrate our proposed technique by an example showing how the "degree" of the principal or governing correlation in a sentence may be determined within certain limits.

No fixed relation exists between the number of correlata and the degree of the governing correlation. But we have established a limited range of possibilities, which becomes more restricted as more of the correlations of the first degree in the given sentence are determined by analysis.

This study has been provisionally suspended because in its present form it is not adaptable to the method of analysis employed by the linguistics sector in the "Multistore" procedure . In itself, however, this line of research has extensive possibilities of practical application. Part of the results of this study, however, have been incorporated in study no. 5 below.

3) RECURRENCE OF A SINGLE EXPLICIT CORRELATOR

When one explicit correlator recurs twice or more consecutively, or when two correlators alternate in a regular sequence, the possible correlational schemes corresponding to such sequences have special characteristics. This research is only at the draft stage, but will be taken up again as soon as there is a possibility of its practical application. The phenomenon is related to YNGVE's "depth-hypothesis".

4) THE GEOMETRY OF CORRELATIONAL SCHEMAS

This work seeks to establish a rigorous formulation of the geometrical representation of correlations which was presented at Elsinore, following the model of Euclidean geometry.

Given the definitions of the constitutive elements, we set up four fundamental postulates, as follows :

1) - every correlator must be preceded and followed by a designatum. This rule corresponds to the fact that in any schema correlators + designata must alternate, and that the first and last items must be designata.

2) - every correlation must have at least one arm which is of unit thickness.

3) - The outer perimeter of any correlational schema, whether partial or complete, is always a rectangle.

4) - Within the perimeter of any correlational schema there must be no empty spaces; the correlations must fill the rectangle completely.

From these postulates 8 theorems have been drawn; these, with their corollaries, constitute the framework of the whole mathematical treatment of correlational schemas. These theorems are not, in the present state of our research, directly applicable to the logical processing of sentences, but we expect to be able to apply them when we know more about the possible interdependences between the items which make up a sentence; interdependances which seem to be determined largely by the type and origin of the ideas themselves.

5) THE ALGEBRA OF CORRELATIONAL SCHEMAS

This study seeks to set up some formulae which obtain between the constituent elements of a schema, to enable the machine to introduce algebraic calculus into the determination of correlational schemas.

The elements involved are the following : the absolute maximum degree and the absolute minimum degree, as a function of N; the influence of the degree of the governing correlation on the possible number of partial correlations, and the influence of the position of the first-degree correlations on the maximum number of the correlations of the intermediate degrees and on the position of the governing correlation.

The eight formulae obtained are these :

$$\alpha^{*}_{max} = N - 1$$

$$\alpha^{*}_{min} = \log_2 N$$

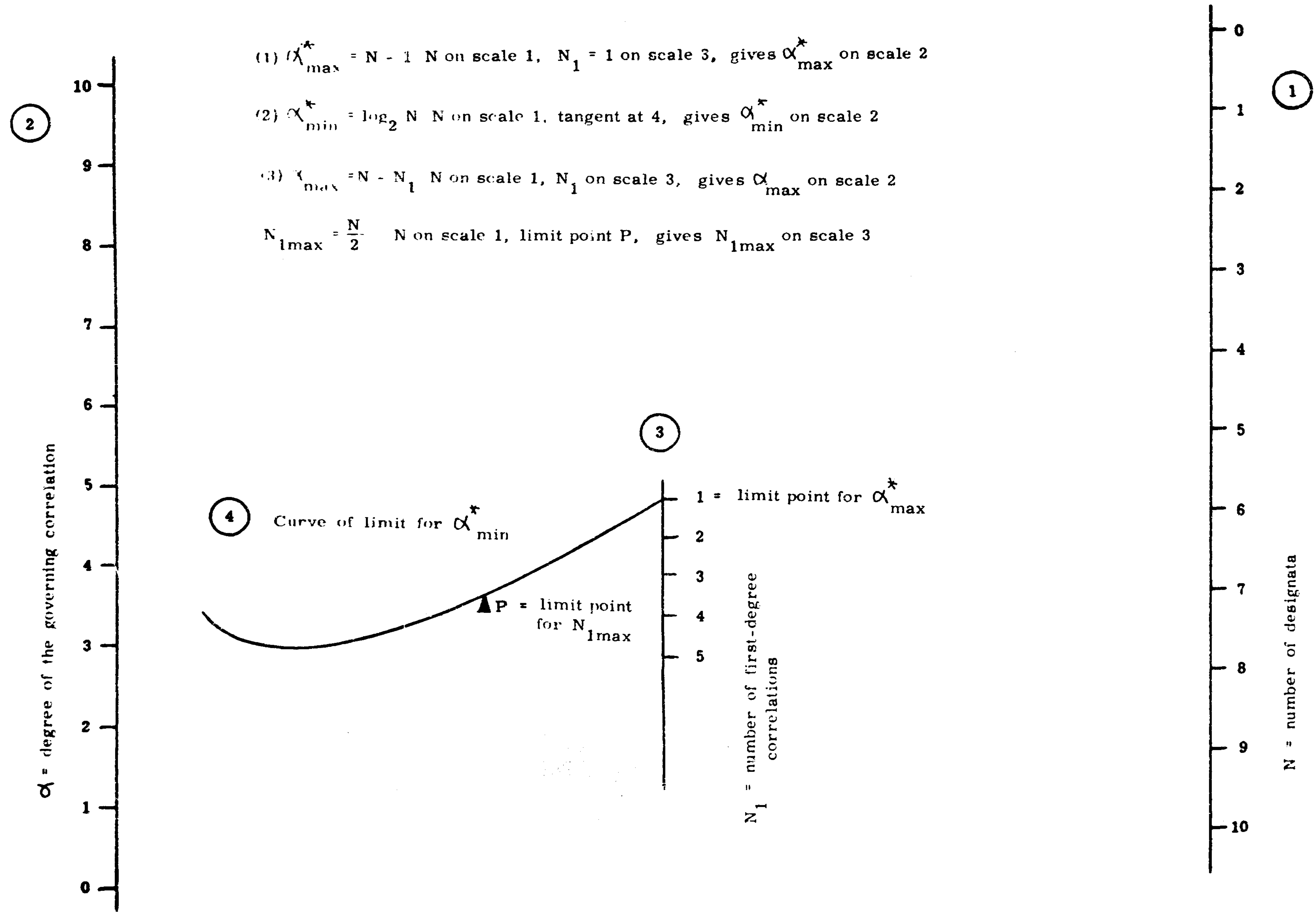
$$\log_2 N \leq \alpha \leq N - 1$$

$$\alpha_{max} = \left(N + m - 1\right) - \sum_{1}^{m} n\, N_n$$

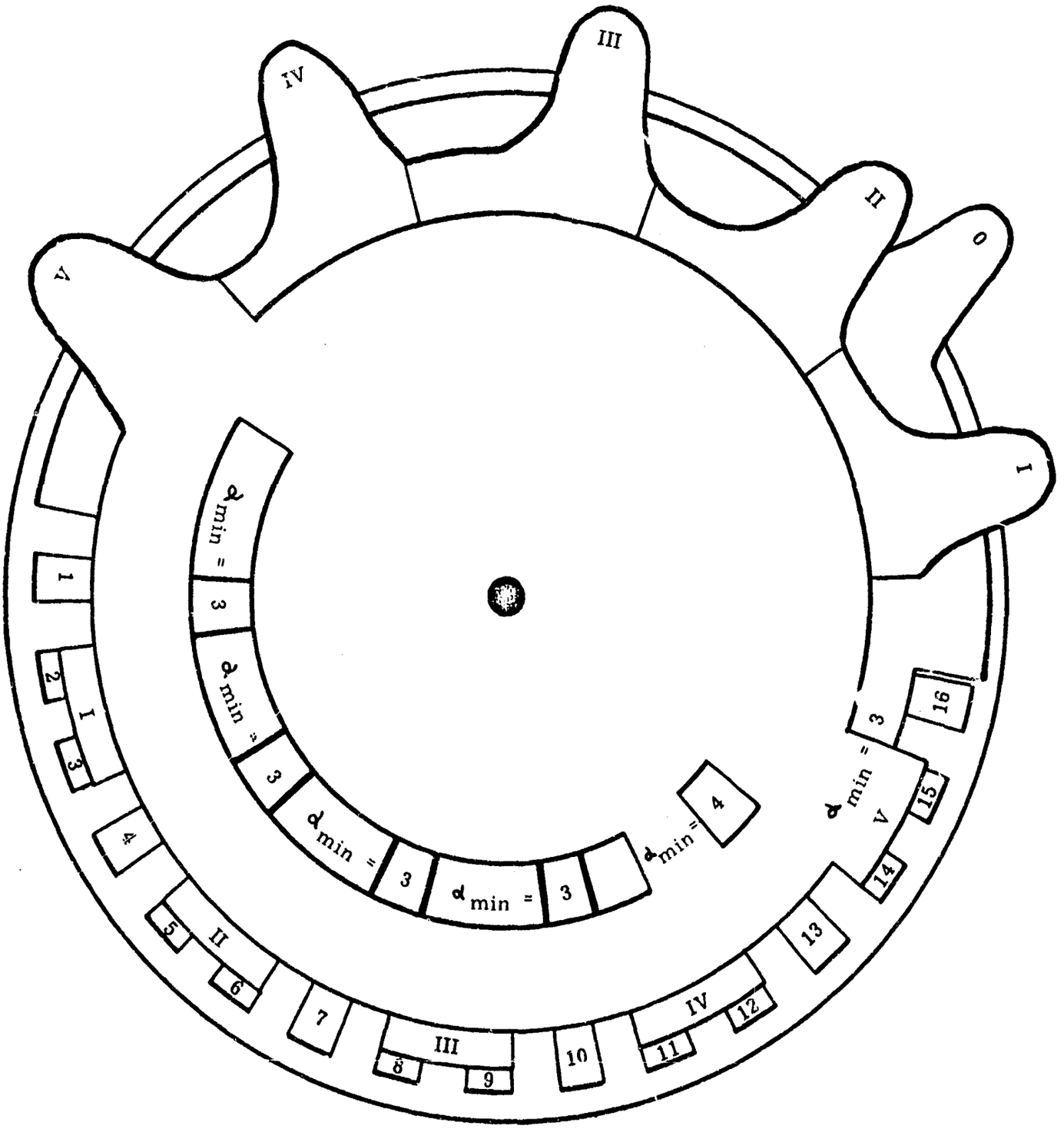
$$N - 1 = N_1 r + N_1 + R\left(\frac{N}{N_1}\right) - 1$$

$$\alpha_{min} = \frac{N}{N_1} - \frac{1}{N_1} R\left(\frac{N}{N_1}\right) + \log_2\left[N_1 + R\left(\frac{N}{N_1}\right)\right] - 1$$

$$\alpha_{min} = \frac{N}{N_1} + \log_2 N_1 - 1$$



(1) $\alpha^*_{max} = N - 1$ N on scale 1, $N_1 = 1$ on scale 3, gives α^*_{max} on scale 2
(2) $\alpha^*_{min} = \log_2 N$ N on scale 1, tangent at 4, gives α^*_{min} on scale 2
(3) $\alpha_{max} = N - N_1$ N on scale 1, N_1 on scale 3, gives α_{max} on scale 2
$N_{1max} = \frac{N}{2}$ N on scale 1, limit point P, gives N_{1max} on scale 3
2
10
9
8
7
6
5
4
3
2
1
0
α = degree of the governing correlation
4
Curve of limit for α^*_{min}
P = limit point for N_{1max}
3
1 = limit point for α^*_{max}
2
3
4
5
N_1 = number of first-degree correlations
1
0
1
2
3
4
5
6
7
8
9
10
N = number of designata



The device comprises several superimposed concentric discs, each of which can be rotated independently. The lowest and largest disc bears on its outer edge a schematic representation of a sequence of designata and correlators, in which the designata are numbered progressively The required number of designata is uncovered by rotating the second disc O

Each of the upper discs - that is, all but the two described above - carries a Roman numeral, and represents one first-degree correlation: the tabs I, II, III, IV can each be moved so as to coincide with any two adjacent designata, thus registering the fact that these two designata are correlated in a first-degree correlation Any disposition of first-degree correlations can thus be represented. In windows in the central area, different values of α_{min} appear, of which we take the highest.

In the example shown, there are 16 designata, with five first-degree correlations connecting the designata 2+3, 5+6, 8+9, 11+12, 14+15. The values of α_{min} are 3, 3, 3, 3, 4, 3; so we take $\alpha_{min} = 4$ as the result.

where:

N = number of designata

N_1 = number of first degree correlations

N_n = number of nth-degree correlations

$R\frac{N}{N_1}$ = result of the division $\frac{N}{N_1}$

α^*_{max} = absolute maximum degree, i. e. maximum value which the governing correlation may have for N designata

α^*_{min} = absolute minimum degree, i. e. minimum value which the governing correlation may have for N designata

α_{max} = maximum degree, i. e. maximum value which the governing correlation may have given certain correlational conditions

α_{min} = minimum degree, i. e. minimum value which the governing correlation may have given certain correlational conditions

We have represented in 16 schemas the possible variations in the degree of the governing correlation in function of its position.

Analysis of the hierarchical structure of mental operations has led to the presentation of certain general principles for the design of a program to calculate:

- the maximum degree of a net of N designata ;
- the minimum degree of a net of N designata ;
- the degree of the governing correlation as a function of the number of partial correlations;
- the degree of the governing correlation as a function of the distribution of the first-degree correlations;
- the maximum number of the correlations of degree K as a function of the position of the governing correlation.

6) NOMOGRAPHIC REPRESENTATION OF THE MATHEMATICAL FORMULAE

The most important formulae, which relate together the constituent elements of the correlational schemas, have been represented in graph form, to facilitate the solution of the formulae themselves.

Ten graphs have been drafted, and two calculating devices have been constructed, using circular scales, for the determination of the degree of the governing correlation. We append a copy of the graph for calculating the maximum and minimum degrees of a schema, and a photograph of the calculating device which yields the degree of the governing correlation as a function of the number and disposition of the first degree correlations.

This work is to be completed, incorporating a classification of mental

items and a hierarchy of relations.

7) THE NUMBER OF POSSIBLE CORRELATIONAL SCHEMAS

The number of alternative schemas corresponding to a correlator-designatum sequence increases exponentially with the number of designata, and quickly reaches enormously high values.

It can be reduced, however, by taking into calculation the following elements:

- the number of first-degree correlations;
- the degree of the governing correlation;
- the number of correlations of intermediate degree;
- the position of the governing correlation.

The results, for schemas comprising up to 16 designata, are shown in a set of tables.

The work can be carried further by taking into consideration the positions as well as the number of the first degree correlations.

We have yet to clarify the influence on the number of possible schemas of the reversibility of certain correlations, and to make a complete study in terms of combinatory calculus, of compatibility among correlations.

The research employs a computer program for the calculation of the number of possible schemas. The program comprised 350 instructions, and was run on an IBM 1401; it calculates the number of different correlational schemas possible with N designata. The inquiry was carried no higher than N=100.

Together with the total number of schemas possible with N designata, in each case the number of alternative half-schemas possible under the right and left arms of the governing correlation, for each of the possible positions of the latter, was also printed out.

8) THE NUMBER OF SCHEMAS AND THE COMBINATORY CALCULUS

The purpose of this study is to check, by the application of combinatory calculus, the number of possible correlational schemas arrived at by other means.

A general formula has been arrived at to calculate :

- how many of the alternative schemas for N designata contain x first-degree correlations;
- how many of the alternative schemas for N designata contain a governing correlation of degree y;
- how many of the alternative schemas for N designata contain a governing correlation of degree y and x first-degree correlations at the same time;
- how many of the alternative schemas for N designata contain x correlations of degree K.

9) RECLASSIFICATION AND THE ALGEBRA OF CLASSES

On the assumption that the algebra of classes, basis of the Boolean calculus of classes, can be applied to the reclassification of correlations, some theoretical schemes of calculation have been sketched. It turns out that in most cases

- this algorithm lends itself to the schematisations which are required in reclassification. This would have to be checked by an extensive examination, prospective and retrospective, of reclassifications - a project too massive to embark upon at present; it would require a full-time worker's attention for two years at least.

10) APPLICATION OF BINARY CALCULUS TO THE CODIFICATION OF CORRELATIONAL SCHEMAS.

The starting-point of this study was the conviction that if a technique could be worked out for representing geometrical schemas in terms of binary numbers, it would become much simpler to handle such schemas in a computer. The study, which is currently under way, runs parallel to that on the hierarchy and the reversibility of correlations.

11) CORRELATIONAL SCHEMAS - HIERARCHY AND REVERSIBILITY

This work, which is being carried on simultaneously at the linguistic and at the geometrical level, has a double purpose:

- on one hand to establish, by way of the study of the reversibility of correlations the hierarchical rules which condition their relative places in a schema;
- on the other to clarify the nature of the structural alternations in a schema necessary in translation from one language to another.

This study has already yielded many results of the greatest interest. Some models have been constructed which clearly display how far the hierarchy of relations remains unaltered in the transition from one language to another, in spite of changes in construction.

Further models are being constructed, programs are being designed for some of the simpler applications of the results.

The mathematical group feels that this is perhaps the most interesting and potentially valuable of the pilot-studies reported here.

12) POLAR REPRESENTATION OF CORRELATIONAL SCHEMAS

We have examined the possibility of a polar representation of correlational schemas. This is of particular interest because it displays the distribution of mental items corresponding to words, in terms of the number of relations to which they are subject.

A more detailed examination could permit a measure of the relative importance of the partial schemas corresponding to the components subject, development and object within the sentence.

This representation will probably be of use, in many cases, in solving the problem of maintained correlations.

13) THE DEGREE OF ELABORATION OF THE PARTIAL SCHEMAS

The general schema of a sentence contains at most five basic components:

- subject;
- verb;
- object;

- time;
- space;

Each of these, however, may be composed either of a single word or of a complete correlational schema, without any alteration to the fundamental structure of the sentence.

An analysis has been made of the operations which may lead to the elaboration of each of the 5 parts of a correlational schema.

The 5 components in their various combinations may each be elaborated according to their particular characteristics.

It seems that the modules according to which these parts are elaborated are characteristic, and definable in terms accessible to the computer.

14) THE DIVISION BETWEEN RELATIONS AND MENTAL ITEMS.

This program has been worked out for the following purposes:

- to separate simple mental relations from the ideas connected by them;
- to examine the sequence in time of relations and of ideas, and their distribution;
- to isolate programming errors on the one hand, and necessary requirements on the other.

In working out this analysis a text was chosen at random, and no account was taken of punctuation.

The program, which was run on an IBM 1410, was difficult to design, though it involved only 500 instructions and a block-diagram. For feeding-in the data, a standard program was used, over and above the 500 instructions of the analysis proper.

15) PROGRAMMING OF THE DIVISION BETWEEN RELATIONS AND IDEAS

This program was worked out for the following purposes :

- to split up compound words ;
- to separate simple mental relations from the ideas connected by them;
- to isolate compound correlators;
- to examine the sequence in time of relations and ideas, and their distribution;
- to recognise errors on the one hand and necessary requirements on the other.

In working out this program, an Italian text was chosen at random;and no account was taken of the value of punctuation. The program would be applicable in any language. The first draft was made in Italian to facilitate the auxiliary operations such as punching, console testing etc.

The program, run on an IBM 1410, was complicated to design, and involved 1650 instructions and a block-diagram. For feeding-in the data a standard program was used, over and above these 1650 instructions.

16a) 1st. PROJECT FOR PROGRAM FOR DIVIDING RELATIONS AND IDEAS

From the programs already worked out, it has become clear that the ambiguity of words which may represent either relations or designata, or which may represent designata of different syntactic functions, needs separate resolution.

While specific rules for the resolution of these ambiguities are not yet available, a program has been simulated, which can be carried out when the rules are complete. The program is at present described in natural language, and comprises 375 non-elementary instructions and two sub-routines, one of which is of indeterminate length in the absence of the complete list of rules. The description is accompanied by a block-diagram.

2nd PROJECT FOR A PROGRAM TO SEPARATE RELATIONS AND IDEAS.

This project differs from that above only in that the recognition of compound correlators and of preliminary idiomatic expressions is performed in the initial phases of the program, and the operation of decomposition is carried out progressively as it becomes necessary in the evolution of the routine.

FURTHER PROJECTS

In the work done so far, various aspects of our problem have been explored some more extensively than others.

We have now to choose between the following lines of research :

1 - to bring to completion as quickly as possible those studies which are at present most advanced, though they may not yet have a practical application (nos. 4, 5, 6, 7, 8).

2 - to develop more thoroughly the analyses which offer the most immediate possibility of practical applications (11, 12, 13).

3 - to continue our work along general exploratory lines, seeking to create as many openings as possible towards wider fields of enquiry.

In this last case, the most useful fields to be explored would seem to be :

- a study of word-frequencies in English texts, to enable us to establish a hierarchy in the ways of handling individual words;
- a classification of ideas in terms of their sources;
- simulation by computer of the mental operations involved in language;
- identification of canonical structures in different languages, in terms of the reversibility of correlations;
- the structure of definitions, on the basis of study 13.

The mathematical group of IDAMI is in fact most inclined to give precedence to the second of the three lines of enquiry described above, while retaining the possibility of following up parts of the others should time be available.

INTRODUCTION TO THE STUDY OF LANGUAGE

The mind, thought, and language (Abstract)

Since language is the partial expression of thought, and since thought is a product of the mind, in order to understand the nature of language we must concern ourselves both with thought and with the mind.

With respect to his consciousness, a man who uses language is an accumulation of experience, a memory, a storehouse of ideas collected over a period of time.

How are these ideas born, and how are they connected among themselves? - that is to say, what is the nature of thought ? This is the most fundamental question. It may seem that it has nothing to do with sentences, with the parts of a discourse, with the consecutio temporum. But in fact, the recognition that the activity of the mind is the origin of every discourse is the only firm basis on which we can begin our study of language. And the sentence is, essentially, the completed coherent expression of an idea.

We must concern ourselves with ideas, then, to discover how, and to what extent, they form a part of language. And to understand the nature of ideas we must examine their origins.

The origin of ideas

We shall begin by asking two questions;one about the subject which produces ideas, and the other about the 'object' with which most ideas are concerned.

1) What is man ? In his capacity as a producer of ideas, man is, first, a set of measuring instruments of great complexity; second, a laboratory which processes, transforms and rearranges the readings from these instruments; and last, as an effect of the exercise of his own will, he is a device which generates fields of values measurable by his own instruments.

2) What is the external world ? The external world is a manifold of fields of values measurable by our measuring instruments.

For example, through our eyes we receive images, which consist of the sum of point-readings of the values of an electromagnetic field; we receive sounds, smells, flavours, colours; we perceive texture and mass, which are all field-values. And having registered these readings we process them in various ways.

A brief history of studies of the mind.

A similar view has been proposed more than once in the past. Socrates, Plato, Aristotle and ther Greek philosophers all examined the question of the origin of ideas.

How does our view of the problem differ from theirs ? The development of this line of thought about the origin of ideas, between the time of the Greeks and the present, can be considered as falling into four sections:

- First, the observations of the empiricists and of the associationists, Locke, Condillac, Hume, Mill, Spencer, Ebbinghaus, to name only the greatest. Note that I speak here only of their observations, not of their conclusions.
- Second, the isolated but extremely important work of Romagnosi on logic, and the qualities of ideas and relations .
- Third, the psychological investigations of William James, immediately followed by those of G. C. Ferrari at Bologna and Wundt in Germany on cerebral operations; investigations carried further by Mc Cullogh.
- Fourth the operative analysis of thought made by Silvio Ceccato, and his attempt to list "correlations".

THE HYPOTHESIS OF THE MIND AS A FACTORY FOR THE PRODUCTION OF IDEAS

This material has been incorporated into a rigorous theoretical structure, representable in a mathematical codification of great flexibility, at IDAMI.

The complex of quantitative readings which our senses give us contains all the data necessary to enable the mind, by processing this complex of readings, to move from the simplest sensory impression to the most sophisticated "abstract" concepts. How is the work of the writers we have cited to be brought to bear on this process ? And what parts of their work will be most useful to us?

From each of them we have taken something, in particular, we have often been able to make use of their initial observations even where we could not accept their conclusions.

From the Greeks, and Socrates in particular, we have adopted the observation that there are certain ideas which are innate, since they have no counterpart in the evidence of our senses: the concepts of causation, analysis etc; but we restrict our acceptance of this observation to the observable fact that only some ideas are innate; those, to be precise, which are determined by our physiological nature.

From the associationists, we have adopted the observation that ideas can arise from the association of other ideas. But we do not follow their conclusions we limit ourselves to the observable fact that only some ideas arise by association.

From the empiricists, we have adopted the observation that ideas derive from sense-data; but whereas they concluded that all ideas derive from sense-data ("nihil est in intellectu quod prius non fuerit in sensu " - Locke) we limit our acceptance to the observable fact that some ideas derive from sense-data.

From Mc Cullogh we have adopted the observation that there are ideas which are derived from the presence of invariants; but we limit our acceptance to the fact that only certain ideas arise from invariants.

From Romagnosi we have adopted especially his study of the characteristics of ideas and of the connections which combine them in thought.

From James, G. C. Ferrari, Wundt, we have adopted the concept that ideas are only formed by means of "cerebral processes".

From Ceccato, lastly, we have adopted :

- all his analysis of the "cerebral processes" which he calls correlations;

- the fact that these correlations are very limited in number;
- his first attempt to enumerate them.

Research at IDAMI

With this material IDAMI has sought to do three things :

- to make a synthesis of it, which provides the framework of an already fairly comprehensive operative scheme;
- to complete this framework by hypothetically filling some of its gaps;
- to identify, in this way, the origins of all the types of ideas which can be isolated in language, tracing them back to the basic operations of the human mind.

MENTAL OPERATIONS

Sensory readings;ideas which are complex at their origin.

Hearing, sight, touch, taste and smell are not the only fields in which the human body can register readings. If we try to lift a table, we are acting as a dynamometer;if we go into a room and feel the warmth of the air, we are acting as a thermometer. When we digest or fail to digest our food we are acting as a highly sensitive chemical plant; and so on.

The measuring-instruments of the human body, like all other instruments, have their fields of application; they can only measure, that is, within well-defined limits; otherwise the machine ceases to register further distinctions, or breaks down, or burns out like a voltmeter or an ammeter.

If our instruments tried to measure temperatures of 1000 degrees, we would be vapourised; if we tried to support weights of fifty thousand tons, our dynamometer would collapse.

If we consider a human being in this way as an elaborate instrument-panel, we can immediately make this first observation - a fundamental one, which gives rise to many others :

the readings taken by the human body never occur singly or in isolation; they are always multiple.

When we look at a room, we see innumerable things simultanously; and at the same time we can feel the temperature, smell tobacco smoke, and hear people talking. All these complex readings are made at once. And these complex readings constitute one source of our ideas; the idea of my house, of an apple from my garden, of the bay of Genoa, and so on. Thus far we agree with Locke.

The analysis of sensory readings : component ideas.

When, by a process of mental analysis, we take from our memory or from our field of attention a group of these sensory readings, and represent to ourselves separately the individual sets of readings which compose it, we create other ideas.

The operative sequence is as follows. In the first place we have the idea of something, when we have registered some of its phisical characteristics and recorded them in our minds as an associated set. We have an idea of this chair, of that person, of that machine - we have an existential idea of these individual objects.

Sensory readings do not only measure static objects, but also things which change in time or in space. They take account of animals moving, of clouds dispersing in the sky; and they go on continuously. From this continuum the mind detaches individual objects, abstracting them from their temporal spatial or causal contexts, and makes component ideas : "eye", "leg", "bough", "star". This, then, is a second source of ideas.

The analysis of sensory readings : mental relations.

This ability of the mind to analyse the continuum of readings presupposes a very important point. In the structure of a properly working mind, which obviously corresponds at least initially to the structure of sensory perceptions, there must exist a priori a mechanism which either by virtue of its structure, or by its way of operating is capable of setting up relations between ideas. Thus, the arrangement of spatial relations recorded by the retina must be conveyed to an organic stratum which determines the way in which these relations are recorded in the memory.

The mind cannot set up relations which its structure does not provide for. From the beginning, that is, the mind comprises in its structure the possibility of setting up the various types of mental relation : spatial, temporal, causal etc.

So that even before the mind receives sensory data, the organic connections which are the key to the operative possibilities of the mind must already be present in its structure, ready to deal with the data when they arrive from the senses. These connections are the organic matrices which form the mental relations (cause-effect, etc).

The perfect organic arrangement, the operativeness of these connections, must necessarily precede sensory experience. Otherwise experience could never be organised into ideas and so into thought. The setting-up of mental relations is an innate function of the mind; but the conscious idea of these relations which we arrive at is a product of reflection.

More precisely, the idea of the relations which the mind sets up in the process of analysis, of synthesis, or of reflection, must be innate implicitly; it becomes explicit if we make it the object of our mental attention.

The mind subsequently employs these same relations when it connects two or more component ideas in a resultant idea; that is, when it synthesises.

To repeat once more: there is, in the structure and the operation of the mind, a set of implicit innate ideas: the ideas of relation. When we say innate, we mean that they consist in the structure and operative possibilities of the mind; in particular of that part of the mind which makes analyses and syntheses and reflects.

This, then, is a third source of ideas; the ideas of the relations which we ourselves set up between things - "before", "after", "above", "and", "without", etc.

Synthesis

The human mind can use these relations to compose complex ideas, drawing their components from the memory. In this way it synthesises new ideas.

Synthesis is a complex operation; we have already gained some insight into its nature, but to discuss it in full is not practicable here. Synthesis is at the root of all technical progress; and produces concepts which are sometimes completely original: the concept of a house, of an arrow, an automobile, an airship. Synthesis constitutes a fourth source of ideas.

Reasoning and abstraction by invariants

The mind has other operative possibilities, however: it can take two ideas, compare them, and study the relations arising from their comparison (ratio) by reasoning; for between two ideas there are both identities and differences.

The identities or invariants, if isolated, form new "abstract" ideas (abs traho), extracted from sensory data and from elsewhere, following for example the sequence: "Fido", "dog", "quadruped", "mammal", "vertebrate", "animal", "living being", "existent".

The process of abstraction begins from ideas derived from sensory data and, by successive reductions in the number of component ideas, arrives at the highest abstraction - the concept of existence, which is common to all ideas. In this we agree entirely with Mc Cullogh.

This, then, is a fifth source of ideas.

Reasoning and concretion by variants.

By concretion we mean the process inverse to that of abstraction.

To the idea of "house", for example, we may add any number of component ideas, and thus arrive at the exact idea of the particular house we wish to construct in its existential reality. The production of ideas by concretion is in fact one of the invariants of the technological process, or the composition of a novel : the Mill on the Floss, Charles's automobile, the White House, etc. This, then, is a sixth source of ideas.

The intelligence, or faculty of choice

The mind can carry out further processing on the ideas which are products of reasoning. By means of "intelligence" (inter lego, I choose between) it can make a choice between two ideas on the basis of the difference between them. We employ reason and intelligence in playing chess, driving a car, working out a project.

From this process we derive the ideas of behaviour, of equivalence, of opposites, etc. This is a seventh source of ideas.

Representation of mental operations

All the operations described above may be carried out unconsciously by the human mind - without its becoming aware that it is analysing, synthesising, reasoning or choosing. But even without being consciously aware of its own processes the mind knows more or less exactly what it does - it has an idea of it. This idea does not derive from our instruments of measurement; it does not come from the external world. It is implicitly innate in us and becomes explicit when we turn our conscious attention to it. And here we can accept completely the findings of Plato.

Here then is an eigth origin of ideas; consciousness of synthesis, analysis, comparison, judgment, etc.

Interpersonal Communication : Codification

We have considered the different ways in which ideas can originate in the human mind. But even the most active mind will not originate, in a lifetime, as many ideas as the memory is able to contain.

But communication is possible between minds;ideas can be conveyed from one mind to another. The contact is not direct - we cannot make the mental structures of one mind directly perceptible to another mind - and the communication which is possible is therefore only partial. We can use, to set up communications with another mind, only one of the sources of ideas discussed above;that is, sensory perception;for this is the only one which puts the mind in relation with the outside world.

In order to transmit thought by way of sensory perceptions, an intermediate codification is necessary, whereby an idea or a complex of ideas can be represented by something accessible to our sense-perceptions. This codification may take diverse forms;but,on the basis of the ability of the human memory to retain two ideas in a fixed association, it involves in every case a fixed association between the idea of a thing and the idea of a signal. It is important that the signals chosen should be clearly distinguishable from other phenomena recorded by the same sense-organs, and that they should fall within the ranges of values which the sense-organs can receive.

The sense which, over millennia, has been most highly trained to distinguish the various signals of a code is the sense of hearing; and concurrently the muscles of the month, tongue and throat have been trained to produce a set of distinguishable sounds which can be combined into code-signals or words . From these signals are then derived the conventional signs of the alphabet;the distinction between sonants and consonants is the basis which permits the wide variety of sound-combinations involved in codification - by - speech.

The basic sonants and consonants are only a few dozen altogether - a few hundred if we can distinguish the intermediate gradations. But this limited set of elements permits of combinations even more numerous than we require for purposes of communication.

Writing is an alternative system of codification; it may be direct, as in the Chinese ideogram where one signal corresponds to one idea, or indirect, as in alphabetic scripts which in effect codify a codification. Both these systems make use of the sense of sight instead of that of hearing.

The codification of language as an expression of thought and a means of communication

Language, then, is a conventional device permitting ideas to be conveyed from one mind to another, ideas which represent not the whole of human thought, but only certain aspects, certain levels of it, which form only a small fraction of the whole.

Thought is composed of an intricate network of ideas and of relations connecting these ideas. To convey any part of this structure from one mind to another, we must observe the following sequence;the mind of the speaker, the

speech-organs of the speaker, the hearing-organs of the recipient, and the mind of the recipient.

The task of conveying thought by means of this sequence can be divided into three parts :

how ideas may be codified;
how relations may be codified;
how the three-dimensional structure of thought can be formulated in terms of a linear, that is, one-dimensional, sequence of signals.

In order to codify ideas, man has developed a signal-code whose significant units are words, to each of which, in general, a given idea corresponds; though the correspondence will not be exactly the same for all the people who use the word.

The connection between an idea, whether simple or complex, and a word, comes about when the idea has a certain stability and frequency of use; the determining factor in the assignation of a word to an idea is habit. The formation of new words is still going on; indeed, it happens now more rapidly than ever before. We may say that every new word conveys a particular idea; but it is not the case that every new idea gives rise to a word, since many ideas, even when they are stable combinations in repeated use, continue to be expressed by a whole sequence of words.

In order to codify relations, besides using the same sorts of signal which codify ideas, we have worked out other techniques; we may convey relations either by an extensive use of inflections as in Latin and Russian; or by using few inflections but embodying in the language strict rules about word-order as in English; or by a mixture of these two methods, in varying proportions as in the Romance languages.

In order to convey in a temporal, one-dimensional sequence, the simultaneous threedimensional structure of thought, the mind must transform this structure into a linear sequence of ideas and relations.

To express a thought, we focus our mental attention on one idea; we then pass to a second idea by way of the relation which connects them, thence by way of a second relation to a third idea, and so on; the attention follows this fixed alternation of ideas and relations until it has fully codified that part of its thought which it wishes to express.

There is a fundamental difference, then, between the structure of thought and that of language.

In thought, every idea is, in general, connected with other ideas by way of more than two mental connections; relations radiate from each idea to connect it with others. In language, on the other hand, one idea is always held between two and only two relations, except in the case of the first and last designata of a sentence, and each relation connects two and only two ideas.

BIBLIOGRAPHY

Our text contains generic references to the works of Euclid, Socrates, Plato, Aristotle, Locke, Condillac, Hume, Mill, Spencer, Ebbinghaus, William James, Boole, Wundt, Ceccato.

Specific references we should like to stress are:

Rusconi, Franco, "Appunti di nomografia", in Rivista di Meccanica, 1st edition, 1962, p. 78ff.

Ferrari, G. C., "Manuale pratico di educazione della volontà", (Concorso L. Brioschi), 1922, p. 11ff.

McCulloch, Warren S., "Why the Mind is in the Head".

Romagnosi, Giandomenico, "Vedute fondamentali sull'arte logica, Reale Accademia d'Italia, Rome, 1936, p. 114ff.
"Che cosa é la mente sana? Indovinello massimo che potrebbe valere poco o niente", Opere, Vol. I.